09/513,999

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450 Alexandria, VA 22313 on February 15, 2005.

Frank C Eisenschenk

Frank C. Eisenschenk, Ph. D., Patent Attorney

REQUEST FOR CERTIFICATE OF CORRECTION UNDER 37 CFR 1.322
Docket No. G-059US02REG
Patent No. 6,783,961

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants	:	Jean-Baptiste Dumas Milne Edwards, Aymeric Duclert, Jean-Yves Giordano
Issued	:	August 31, 2004
Patent No.	:	6,783,961 B1
For	:	Expressed Sequence Tags and Encoded Human Proteins

Mail Stop Certificate of Corrections Branch
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

REQUEST FOR CERTIFICATE OF CORRECTION
UNDER 37 CFR 1.322 (OFFICE MISTAKE)

Sir:

A Certificate of Correction (in duplicate) for the above-identified patent has been prepared and is attached hereto.

In the left-hand column below is the column and line number where errors occurred in the patent. In the right-hand column is the page and line number in the application where the correct information appears.

Patent Reads:	**Application Reads:**
Column 1, line 27:	Page 1, line 19:
"kilobases (cb)"	-- kilobases (kb) --
Column 2, line 31:	Page 3, line 3:
"fraes"	-- frames --

Column 2, line 31:	Page 3, line 3:
"Biotech"	-- Biochem --
Column 15, line 15:	Page 22, line 3:
"code of SEQ D"	-- code of SEQ ID --
Column 16, line 24:	Page 23, line 25:
"The "a" in the"	-- The "/" in the --
Column 16, line 46:	Page 24, line 8:
"The "P" in the"	-- The "/" in the --
Column 18, line 8:	Page 26, line 12:
"fingal"	-- fungal --
Column 18, line 55:	Page 27, line 13:
"(Strajagene)"	-- (Stratagene) --
Column 21, line 33:	Page 31, line 11:
"subunit a and "	-- subunit α and --
Column 22, line 22:	Page 32, line 18:
"convex"	-- connex --
Column 22, line 25:	Page 32, line 21:
"convex"	-- connvex --
Column 23, line 58:	Page 34, line 25:
"(SEQ IE NOs:3721-4100)"	-- (SEQ ID NOs:3721-4100) --
Column 27, line 33:	Page 39, line 30:
"biotin-UJT"	-- biotin-UTP --

Column 27, line 34:	Page 39, line 30:
"DIG-UITP"	-- DIG-UTP --
Column 28, line 49:	Page 41, line 23:
"270:467470"	-- 270:467-470 --
Column 29, line 11:	Page 42, line 12:
"Pictu"	-- Pietu --
Column 31, line 8:	Page 45, line 7:
"bioinfomatics"	-- bioinformatics --
Column 32, line 29:	Page 47, line 7:
"if the fall"	-- if the full --
Column 36, line 31:	Page 53, line 9:
"Examples 14"	-- Examples 1-4 --
Column 37, lines 16 and 17:	Page 54, lines 16 and 17:
"(fraction G+C) 0.63% formamide-(600/N)"	-- (fraction G+C)-(0.63% formamide-(600/N)--
Column 37, line 53:	Page 55, line 10:
"Lower Dewees"	-- Lower Degrees --
Column 39, line 48:	Page 58, line 12:
"Cutoff Score, Gap"	-- Cutoff Score=1, Gap --
Column 40, line 45:	Page 59, line 24:
"Mismatch Penaltyl,"	-- Mismatch Penalty=1, --

Column 43, lines 8 and 9:	Page 63, line 9:
"78-8-3-E6-CL1_1C)"	-- 78-8-3-E6-CL0_1C) --
Column 43, lines 50 and 51:	Page 64, line 6:
"108-004-05-0D10-FLC)"	-- 108-004-05-0-D10-FLC) --
Column 49, line 15:	Page 71, line 30:
"DAIG"	-- DA1G --
Column 51, line 59:	Page 75, line 22:
"blocling"	-- blocking --
Column 52, line 44:	Page 76, line 26:
"CTLA41g"	-- CTLA4Ig --
Column 52, line 49:	Page 76, line 29:
"pp. 846847"	-- pp. 846-847--
Column 53, line 60:	Page 78, line 20:
"IIα chain"	-- II β chain --
Column 54, line 16:	Page 79, line 5:
"Hematogoiesis"	-- Hematopoiesis --
Column 56, line 53:	Page 82, line 25:
"neuropatuies"	-- neuropathies --
Column 57, line 58:	Page 84, line 11:
"inhibin a"	-- inhibin α --
Column 61, line 45:	Page 89, line 25:
"arc inserted"	-- are inserted --

Column 64, line 30:	Page 93, line 27:
"pug/ml."	-- μg/ml. --
Column 65, line 13:	Page 94, line 30:
"semifuantitatively"	-- semi-quantitatively --
Column 69, line 25:	Page 100, line 30:
"with 2 using"	-- with P^{32} using --
Column 69, line 26:	Page 100, line 30:
"kinase Pharmacia)."	-- kinase (Pharmacia). --
Column 70, line 44:	Page 102, line 24:
"Ma ping"	-- Mapping --
Column 71, line 11:	Page 103, lines 15 and 16:
"region are surrounding"	-- region surrounding --
Column 71, line 14:	Page 103, line 19:
"14:574584"	-- 14:574-584 --
Column 73, line 35:	Page 106, line 24:
"Acids Positional"	-- Acids, Positional --
Column 73, line 57:	Page 107, line 6:
"posifional"	-- positional --
Column 76, line 13:	Page 110, line 17:
"DEAE-Dextrar"	-- DEAE-Dextran --

Column 76, lines 63 and 64:	Page 111, lines 17 and 18:
"fragments a of"	-- fragments of --
Column 77, line 61:	Page 112, line 30:
"semiualitative"	-- semi-qualitative --
Column 78, line 15:	Page 113, line 13:
"rhodailne"	-- rhodamine --
Column 78, line 25:	Page 113, line 19:
"example 1251"	-- example ^{125}I --
Column 80, line 53:	Page 117, line 6:
"CLAB"	-- CL-4B --
Column 82, line 18:	Page 119, line 9:
"DNTP"	-- dNTP --
Column 86, line 28:	Page 125, line 5:
"gel shill"	-- gel shift --
Column 87, lines 20 and 21:	Page 126, lines 12 and 13:
"acids are A complementary"	-- acids are complementary --
Column 97, line 14:	Page 140, line 20:
"EMBUSwissprotein"	-- EMBL/Swissprotein --.

A true and correct copy of pages 1, 3, 22, 23, 24, 26, 27, 31, 32, 34, 39, 41, 42, 45, 47, 53, 54, 55, 58, 59, 63, 64, 71, 75, 76, 78, 79, 82, 84, 89, 93, 94, 100, 102, 103, 106, 107, 110, 111, 112, 113, 117, 119, 125, 126, and 140 of the specification as filed which supports Applicants' assertion of the errors on the part of the Patent Office accompanies this Certificate of Correction.

Approval of the Certificate of Correction is respectfully requested.

Respectfully submitted,

Frank C. Eisenschenk, Ph. D.
Patent Attorney
Registration No. 45,332
Phone No.: 352-375-8100
Fax No.: 352-372-5800
Address: P.O. Box 142950
Gainesville, FL 32614-2950

FCE/amh

Attachments: Certificate of Correction; copy of pages 1, 3, 22, 23, 24, 26, 27, 31, 32, 34, 39, 41, 42, 45, 47, 53, 54, 55, 58, 59, 63, 64, 71, 75, 76, 78, 79, 82, 84, 89, 93, 94, 100, 102, 103, 106, 107, 110, 111, 112, 113, 117, 119, 125, 126, and 140 of the specification

EXPRESSED SEQUENCE TAGS AND ENCODED HUMAN PROTEINS

Related Application Data

This application claims priority from U.S. Provisional Patent Application Serial No. 60/122,487, filed February 26, 1999, the disclosure of which is incorporated herein by reference in its entirety.

Background of the Invention

The estimated 50,000-100,000 genes scattered along the human chromosomes offer tremendous promise for the understanding, diagnosis, and treatment of human diseases. In addition, probes capable of specifically hybridizing to loci distributed throughout the human genome find applications in the construction of high resolution chromosome maps and in the identification of individuals.

In the past, the characterization of even a single human gene was a painstaking process, requiring years of effort. Recent developments in the areas of cloning vectors, DNA sequencing, and computer technology have merged to greatly accelerate the rate at which human genes can be isolated, sequenced, mapped, and characterized. Cloning vectors such as yeast artificial chromosomes (YACs) and bacterial artificial chromosomes (BACs) are able to accept DNA inserts ranging from 300 to 1000 kilobases (kb) or 100-400 kb in length respectively, thereby facilitating the manipulation and ordering of DNA sequences distributed over great distances on the human chromosomes. Automated DNA sequencing machines permit the rapid sequencing of human genes. Bioinformatics software enables the comparison of nucleic acid and protein sequences, thereby assisting in the characterization of human gene products.

Currently, two different approaches are being pursued for identifying and characterizing the genes distributed along the human genome. In one approach, large fragments of genomic DNA are isolated, cloned, and sequenced. Potential open reading frames in these genomic sequences are identified using bioinformatics software. However, this approach entails sequencing large stretches of human DNA which do not encode proteins in order to find the protein encoding sequences scattered throughout the genome. In addition to requiring extensive sequencing, the bioinformatics software may mischaracterize the genomic sequences obtained. Thus, the software may produce false

Severson *et al., Eur J Biochem* **229**:426-32, 1995) or secondary structures such as IREs (Rouault and Klausner, *Curr Top Cell Regul* **35**:1-19, 1997), and (iii) upstream open reading frames or uORFs (Geballe and Morris, *Trends Biochem Sci* **19**:159-64, 1994). Thus, regulation of gene expression may be achieved through the use of alternative 5'UTRs. For instance, the translation of the tissue inhibitor of metalloprotease mRNA is enhanced in mitogenically activated cells through modification of the start codon of an uORF in its 5'UTR using an alternative promoter (Waterhouse *et al, J Biol Chem.* **265**:5585-9. 1990). Furthermore, modification of 5'UTR through mutation, insertion or translocation events may even be implied in pathogenesis. For instance, the fragile X syndrome, the most common cause of inherited mental retardation, is partly due to an insertion of multiple CGG trinucleotides in the 5'UTR of the fragile X mRNA resulting in the inhibition of protein synthesis via ribosome stalling (Feng *et al, Science* **268**:731-4, 1995). An aberrant mutation in regions of the 5'UTR known to inhibit translation of the proto-oncogene *c-myc* was shown to result in upregulation of C-myc protein levels in cells derived from patients with multiple myelomas (Willis *et al, Curr Top Microbiol Immunol* **224**:269-76, 1997). However, the use of oligo-dT primed cDNA libraries does not allow the isolation of complete 5'UTRs since such obtained incomplete sequences may not include the first exon of the mRNA, particularly in situations where the first exon is short. Furthermore, they may not include some exons, often short ones, which are located upstream of splicing sites. Thus, there is a need to obtain sequences derived from the 5' ends of mRNAs.

While many sequences derived from human chromosomes have practical applications, approaches based on the identification and characterization of those chromosomal sequences which encode a protein product are particularly relevant to diagnostic and therapeutic uses. In some instances, the sequences used in such therapeutic or diagnostic techniques may be sequences which encode proteins which are secreted from the cell in which they are synthesized, as well as the secreted proteins themselves, are particularly valuable as potential therapeutic agents. Such proteins are often involved in cell to cell communication and may be responsible for producing a clinically relevant response in their target cells. In fact, several secretory proteins, including tissue plasminogen activator, G-CSF, GM-CSF, erythropoietin, human growth hormone, insulin, interferon-α, interferon-β, interferon-γ, and interleukin-2, are currently in clinical use.

Another embodiment of the present invention is a method for identifying a feature in a sequence selected from the group consisting of a nucleic acid code of SEQID NOs: 24-4100 and 8178-36681 and a polypeptide code of SEQ ID NOs: 4101-8177 comprising the steps of reading said sequence through the use of a computer program which identifies features in sequences and identifying features in said sequence with said computer program.

Another embodiment of the present invention is a vector comprising a nucleic acid according to any one of the nucleic acids described above.

Another embodiment of the present invention is a host cell containing the above vector.

Another embodiment of the present invention is a method of making any of the nucleic acids described above comprising the steps of introducing said nucleic acid into a host cell such that said nucleic acid is present in multiple copies in each host cell and isolating said nucleic acid from said host cell.

Another embodiment of the present invention is a method of making a nucleic acid of any of the nucleic acids described above comprising the step of sequentially linking together the nucleotides in said nucleic acids.

Another embodiment of the present invention is a method of making any of the polypeptides described above wherein said polypeptides is 150 amino acids in length or less comprising the step of sequentially linking together the amino acids in said polypeptide.

Another embodiment of the present invention is a method of making any of the polypeptides described above wherein said polypeptides is 120 amino acids in length or less comprising the step of sequentially linking together the amino acids in said polypeptides.

Brief Description of the Sequence Listing

SEQ ID NOs: 1, 3, 5, 7, 9, 11, and 13 are full-length cDNAs prepared using the methods described herein.

SEQ ID NOs: 2, 4, 6, 8, 10, 12, and 14 are the polypeptides encoded by the nucleic acids of SEQ ID NOs: 1, 3, 5, 7, 9, 11, and 13.

SEQ ID NOs: 15, 16, 18, 19, 21 and 22 are primers whose use is described in the specification.

SEQ ID NOs: 17, 20, and 23 are the sequences of nucleic acids containing transcription factor binding sites which were obtained as described below.

SEQ ID NOs: 24-652 are nucleic acids having an incomplete ORF which encodes a signal peptide. As used herein, an "incomplete ORF" is an open reading frame in which a start codon has been identified but no stop codon has been identified. The locations of the incomplete ORFs and sequences encoding signal peptides are listed in the accompanying Sequence Listing. In addition, the von Heijne score of the signal peptide computed as described below is listed as the "score" in the accompanying Sequence Listing. The sequence of the signal-peptide is listed as "seq" in the accompanying Sequence Listing. The "/" in the signal peptide sequence indicates the location where proteolytic cleavage of the signal peptide occurs to generate a mature protein.

SEQ ID NOs: 653-3720 are nucleic acids having an incomplete ORF in which no sequence encoding a signal peptide has been identified to date. However, it remains possible that subsequent analysis will identify a sequence encoding a signal peptide in these nucleic acids. The locations of the incomplete ORFs are listed in the accompanying Sequence Listing.

SEQ ID NOs: 3721-3811 are nucleic acids having a complete ORF which encodes a signal peptide. As used herein, a "complete ORF" is an open reading frame in which a start codon and a stop codon have been identified. The locations of the complete ORFs and sequences encoding signal peptides are listed in the accompanying Sequence Listing. In addition, the von Heijne score of the signal peptide computed as described below is listed as the "score" in the accompanying Sequence Listing. The sequence of the signal-peptide is listed as "seq" in the accompanying Sequence Listing. The "/" in the signal peptide sequence indicates the location where proteolytic cleavage of the signal peptide occurs to generate a mature protein.

SEQ ID NOs: 3812-4100 are nucleic acids having a complete ORF in which no sequence encoding a signal peptide has been identified to date. However, it remains possible that subsequent analysis will identify a sequence encoding a signal peptide in these nucleic acids. The locations of the complete ORFs are listed in the accompanying Sequence Listing.

SEQ ID NOs: 4101-4729 are "incomplete polypeptide sequences" which include a signal peptide. Incomplete polypeptide sequences" are polypeptide sequences encoded by nucleic acids in which a start codon has been identified but no stop codon has been identified. These polypeptides are encoded by the nucleic acids of SEQ ID NOs: 24-652. The location of the signal peptide is listed in the accompanying Sequence Listing. In addition, the von Heijne score of the signal peptide computed as described below is listed as the "score" in the accompanying Sequence Listing. The sequence of the signal-peptide is listed as "seq" in the accompanying Sequence Listing. The "/" in the signal peptide sequence indicates the location where proteolytic cleavage of the signal peptide occurs to generate a mature protein.

SEQ ID NOs: 4730-7797 are incomplete polypeptide sequences in which no signal peptide has been identified to date. However, it remains possible that subsequent analysis will identify a signal peptide in these polypeptides. These polypeptides are encoded by the nucleic acids of SEQ ID NOs: 653-3720.

SEQ ID NOs: 7798-7888 are "complete polypeptide sequences" which include a signal peptide. "Complete polypeptide sequences" are polypeptide sequences encoded by nucleic acids in which a start codon and a stop codon have been identified. These polypeptides are encoded by the nucleic acids of SEQ ID NOs: 3721-3811. The location of the signal peptide is listed in the accompanying Sequence Listing. In addition, the von Heijne score of the signal peptide computed as described below is listed as the "score" in the accompanying Sequence Listing. The sequence of the signal-peptide is listed as "seq" in the accompanying Sequence Listing. The "/" in the signal peptide sequence indicates the location where proteolytic cleavage of the signal peptide occurs to generate a mature protein.

SEQ ID NOs: 7889-8177 are complete polypeptide sequences in which no signal peptide has been identified to date. However, it remains possible that subsequent analysis will identify a signal peptide in these polypeptides. These polypeptides are encoded by the nucleic acids of SEQ ID NOs.:3812-4100.

SEQ ID NOs: 8178-36681 are nucleic acid sequences in which no open reading frame has been conclusively identified to date. However, it remains possible subsequent analysis will identify an open reading frame in these nucleic acids.

Sacchi, *Analytical Biochemistry* **162**:156-159, 1987). PolyA$^+$ RNA was isolated from total RNA (LABIMO) by two passes of oligo dT chromatography, as described by Aviv and Leder., *Proc. Natl. Acad. Sci. USA* **69**:1408-1412, 1972) in order to eliminate ribosomal RNA.

The quality and the integrity of the polyA+ RNAs were checked. Northern blots hybridized with a globin probe were used to confirm that the mRNAs were not degraded. Contamination of the polyA$^+$ mRNAs by ribosomal sequences was checked using Northern blots and a probe derived from the sequence of the 28S rRNA. Preparations of mRNAs with less than 5% of rRNAs were used in library construction. To avoid constructing libraries with RNAs contaminated by exogenous sequences (prokaryotic or fungal), the presence of bacterial 16S ribosomal sequences or of two highly expressed fungal mRNAs was examined using PCR.

Following preparation of the mRNAs from various tissues an oligonucleotide tag was specifically attached to the caps at the 5′ ends of the mRNAs. The oligonucleotide tag had an EcoRI site therein to facilitate later cloning procedures. Following attachment of the oligonucleotide tag to the mRNA, the integrity of the mRNA was examined by performing a Northern blot with 200 to 500 ng of mRNA using a probe complementary to the oligonucleotide tag before performing the first strand synthesis described in Example 2.

EXAMPLE 2

cDNA Synthesis Using mRNA Templates Having Intact 5′ Ends

For the mRNAs joined to oligonucleotide tags, first strand cDNA synthesis was performed using a reverse transcriptase with random nonamers as primers. In order to protect internal EcoRI sites in the cDNA from digestion at later steps in the procedure, methylated dCTP was used for first strand synthesis. After removal of RNA by an alkaline hydrolysis, the first strand of cDNA was precipitated using isopropanol in order to eliminate residual primers.

The second strand of the cDNA was synthesized with a Klenow fragment using a primer corresponding to the 5′end of the ligated oligonucleotide. Methylated dCTP was also used for second strand synthesis in order to protect internal EcoRI sites in the cDNA from digestion during the cloning process.

Following cDNA synthesis, the cDNAs were cloned into pBlueScript as described in Example 3 below.

EXAMPLE 3

Cloning of cDNAs derived from mRNA with intact 5' ends into BlueScript

Following second strand synthesis, the ends of the cDNA were blunted with T4 DNA polymerase (Biolabs) and the cDNA was digested with EcoRI. Since methylated dCTP was used during cDNA synthesis, the EcoRI site present in the tag was the only hemi-methylated site, hence the only site susceptible to EcoRI digestion. The cDNA was then size fractionated using exclusion chromatography (AcA, Biosepra) and fractions corresponding to cDNAs of more than 150 bp were pooled and ethanol precipitated. The cDNA was directionally cloned into the SmaI and EcoRI ends of the phagemid pBlueScript vector (Stratagene). The ligation mixture was electroporated into bacteria and propagated under appropriate antibiotic selection.

Clones containing the oligonucleotide tag attached were then selected as described in Example 4 below.

EXAMPLE 4

Selection of Clones Having the Oligonucleotide Tag Attached Thereto

The plasmid DNAs containing 5' EST libraries made as described above were purified (Qiagen). A positive selection of the tagged clones was performed as follows. Briefly, in this selection procedure, the plasmid DNA was converted to single stranded DNA using gene II endonuclease of the phage F1 in combination with an exonuclease (Chang *et al.*, **Gene 127**:95-8, 1993) such as exonuclease III or T7 gene 6 exonuclease. The resulting single stranded DNA was then purified using paramagnetic beads as described by Fry *et al.*, *Biotechniques*, **13**: 124-131, 1992. In this procedure, the single stranded DNA was hybridized with a biotinylated oligonucleotide having a sequence corresponding to the 3' end of the oligonucleotide tag. Clones including a sequence complementary to the biotinylated oligonucleotide were captured by incubation with streptavidin coated magnetic beads followed by magnetic selection. After capture of the positive clones, the plasmid DNA was released from the magnetic beads and converted into double stranded DNA using a DNA polymerase such as the Thermosequenase

EST sequences were ignored to avoid the inclusion of spurious cloning sites in the analysis of sequencing accuracy.

This analysis revealed that the sequences incorporated in the NETGENE™ database had an accuracy of more than 99.5%.

EXAMPLE 8

Determination of Efficiency of 5' EST Selection

To determine the efficiency at which the above selection procedures isolated 5' ESTs which included sequences close to the 5' end of the mRNAs from which they derived, the sequences of the ends of the 5' ESTs derived from the elongation factor 1 subunit α and ferritin heavy chain genes were compared to the known cDNA sequences of these genes. Since the transcription start sites of both genes are well characterized, they may be used to determine the percentage of derived 5' ESTs which included the authentic transcription start sites.

For both genes, more than 95% of the obtained 5' ESTs actually included sequences close to or upstream of the 5' end of the corresponding mRNAs.

To extend the analysis of the reliability of the procedures for isolating 5' ESTs from ESTs in the NetGene™ database, a similar analysis was conducted using a database composed of human mRNA sequences extracted from GenBank database release 97 for comparison. The 5' ends of more than 85% of 5' ESTs derived from mRNAs included in the GeneBank database were located close to the 5' ends of the known sequence. As some of the mRNA sequences available in the GenBank database are deduced from genomic sequences, a 5' end matching with these sequences will be counted as an internal match. Thus, the method used here underestimates the yield of ESTs including the authentic 5' ends of their corresponding mRNAs.

EXAMPLE 9

Clustering of the 5' ESTs

Since the cDNA libraries made above include multiple 5' ESTs derived from the same mRNA, overlapping 5'ESTs may be assembled into continuous sequences. The following method (see Figure 1) describes how to efficiently cluster 5'ESTs in order to

yield not only consensus 5'EST sequences for mRNAs derived from different genes but also consensus 5'EST sequences for different mRNAs, so called variants, transcribed from the same gene such as alternatively spliced mRNAs. This clustering was performed on a set of NetGene™ 5'ESTs sequences following elimination of endogenous contaminants, elimination of uninformative sequences and masking of repeats.

The whole set of sequences was first partitioned into smaller sets, so-called clusters, containing sequences exhibiting perfect matches with each other on a given length. Such clusters contain 5'ESTs derived from a small number of different genes. Some 5'EST sequences were not clustered using this approach either because they were not homologous to any other sequence or because the homology was not properly detected. To overcome this problem, sequences not clustered, so called singletons, may be compared to the consensus contigated ESTs obtained later on and, if necessary, included in the appropriate clusters and used to compute other consensus contigated ESTs.

Thereafter, all variants of a given gene were identified in each cluster as follows. Overlapping sequences inside a given cluster were figured as oriented graphs where each sequence was a node and each overlap an edge. Then, the different genes contained within a single graph which were represented by different connex components were identified and isolated from each other. Subsequently, the different variants of a same gene were isolated using an algorithm based on the detection of forks within a connex component. If desired, the consensus contigated EST sequences may be verified by identifying clones in nucleic acid samples derived from biological tissues, such as cDNA libraries, which hybridize to the probes based on the sequences of the consensus contigated ESTs and sequencing them.

Overlapping 5'EST sequences belonging to the same variant as well as included 5'EST sequences belonging to the same cluster were then contigated and consensus contigated 5'EST sequences were generated for each variant. Some of the obtained consensus contigated 5'EST sequences were incomplete due to the fact that only included and overlapping 5'EST sequences were considered to isolate genes and due to the algorithm developed to find variants. These variant consensus contigated 5'EST sequences were extended as follows. Variants transcribed from the same gene were compared pairwise and the 5' EST consensus sequences that were incomplete either in

EXAMPLE 11

Sequence Analysis

Application of the clustering method described in Example 9 to a selected set of 126,735 NetGene™ 5'ESTs free from endogenous contaminants and uninformative sequences yielded 9490 consensus assembled 5'EST sequences or variants for a total of 8037 genes clustered representing 98,973 individual 5'ESTs. One of them which contained 21,138 sequences and was shown to contain chimeras thanks to comparison to public sequences was removed from further analysis.

Both non clustered 5'ESTs, *i.e.* singletons, and consensus contigated 5'ESTs were then compared to already known sequences as follows. Those sequences matching human mRNA sequences were eliminated from further analysis. Then, following masking of repeats those sequences matching sequences that have already been discovered by the inventors, namely sequences exhibiting more than 90% homology over stretches longer than 40 nucleotides using BLAST2N with overhangs shorter than 10 nucleotides, were removed from further consideration. The final set represents the sequences of the invention (SEQ ID NOs:24-4100 and 8178-36681), *i.e.,* 7609 consensus contigated 5'EST from 6398 clusters containing 31,267 5'ESTs and 24, 972 singletons.

Of the 6398 obtained clusters, 658 were shown to be multivariant, *i.e.* to contain several variants of the same gene. Table I gives for each of the multivariant clusters named by its internal reference (first column), the list of the consensus sequences of all variants, each variant being represented by a different SEQ ID NO.

Subsequently, the most probable open reading frame was determined, as described in Example 10, for all sequences of the invention. 3,697 5'ESTs (SEQ ID NOs:24-3720) encoding incomplete ORFs (SEQ ID NOs:4101-7797) of at least 50 amino acid long were found. In addition, 380 5'ESTs (SEQ ID NOs:3721-4100) encoding complete ORFs (SEQ ID NOs:7798-8177) of at least 100 amino acids were found.

The nucleotide sequences of the SEQ ID NOs: 24-4100 and 8178-36681 and the amino acid sequences encoded by SEQ ID NOs: 24-4100 (i.e. amino acid sequences of SEQ ID NOs: 4101-8177) are provided in the appended sequence listing. Some of the amino acid sequences may contain "Xaa" designators. These "Xaa" designators indicate either (1) a residue which cannot be identified because of nucleotide sequence ambiguity

Furthermore, 5' ESTs and consensus contigated 5' ESTs whose corresponding mRNAs are associated with disease states may also be identified. For example, a particular disease may result from the lack of expression, over expression, or under expression of a mRNA corresponding to a 5' EST or consensus contigated 5' EST. By comparing mRNA expression patterns and quantities in samples taken from healthy individuals with those from individuals suffering from a particular disease, 5' ESTs or consensus contigated 5' ESTs responsible for the disease may be identified.

It will be appreciated that the results of the above characterization procedures for 5' ESTs and consensus contigated 5' ESTs also apply to extended cDNAs (obtainable as described below) which contain sequences adjacent to the 5' ESTs and consensus contigated 5' ESTs. It will also be appreciated that if desired, characterization may be delayed until extended cDNAs have been obtained rather than characterizing the 5' ESTs or consensus contigated 5' ESTs themselves.

EXAMPLE 16

Evaluation of Expression Levels and Patterns of mRNAs Corresponding to EST-Related Nucleic Acids

Expression levels and patterns of mRNAs corresponding to EST-related nucleic acids may be analyzed by solution hybridization with long probes as described in International Patent Application No. WO 97/05277, the entire contents of which are hereby incorporated by reference. Briefly, an EST-related nucleic acid, fragment of an EST related nucleic acid, positional segment of an EST-related nucleic acid, or fragment of a positional segment of an EST-related nucleic acid corresponding to the gene encoding the mRNA to be characterized is inserted at a cloning site immediately downstream of a bacteriophage (T3, T7 or SP6) RNA polymerase promoter to produce antisense RNA. Preferably, the EST-related nucleic acid, fragment of an EST related nucleic acid, positional segment of an EST-related nucleic acid, or fragment of a positional segment of an EST-related nucleic acid is 100 or more nucleotides in length. The plasmid is linearized and transcribed in the presence of ribonucleotides comprising modified ribonucleotides (i.e. biotin-UTP and DIG-UTP). An excess of this doubly labeled RNA is hybridized in solution with mRNA isolated from cells or tissues of interest. The hybridizations are performed under standard stringent conditions (40-50°C for 16 hours in an 80%

extended cDNAs are expressed in the cell, tissue, organism, or other source of nucleic acids from which the tags were derived. In this way, the expression pattern of the 5' ESTs, contigated consensus 5' ESTs, or extended cDNAs in the cell, tissue, organism, or other source of nucleic acids is obtained.

Quantitative analysis of gene expression may also be performed using arrays. As used herein, the term array means a one dimensional, two dimensional, or multidimensional arrangement of EST-related nucleic acids, fragments of EST related nucleic acids, positional segments EST-related nucleic acids, or fragments of positional segments of EST-related nucleic acids. Preferably, the EST-related nucleic acids, fragments of EST related nucleic acids, positional segments EST-related nucleic acids, or fragments of positional segments of EST-related nucleic acids are at least 15 nucleotides in length. More preferably, the EST-related nucleic acids, fragments of EST related nucleic acids, positional segments EST-related nucleic acids, or fragments of positional segments of EST-related nucleic acids are at least 100 nucleotide long. More preferably, the fragments are more than 100 nucleotides in length. In some embodiments, the EST-related nucleic acids, fragments of EST related nucleic acids, positional segments EST-related nucleic acids, or fragments of positional segments of EST-related nucleic acids may be more than 500 nucleotides long.

For example, quantitative analysis of gene expression may be performed with EST-related nucleic acids, fragments of EST related nucleic acids, positional segments EST-related nucleic acids, or fragments of positional segments of EST-related nucleic acids in a complementary DNA microarray as described by Schena *et al.* (*Science* **270**:467-470, 1995; *Proc. Natl. Acad. Sci. U.S.A.* **93**:10614-10619, 1996). EST-related nucleic acids, fragments of EST related nucleic acids, positional segments EST-related nucleic acids, or fragments of positional segments of EST-related nucleic acids are amplified by PCR and arrayed from 96-well microtiter plates onto silylated microscope slides using high-speed robotics. Printed arrays are incubated in a humid chamber to allow rehydration of the array elements and rinsed, once in 0.2% SDS for 1 min, twice in water for 1 min and once for 5 min in sodium borohydride solution. The arrays are submerged in water for 2 min at 95°C, transferred into 0.2% SDS for 1 min, rinsed twice with water, air dried and stored in the dark at 25°C.

Cell or tissue mRNA is isolated or commercially obtained and probes are prepared by a single round of reverse transcription. Probes are hybridized to 1 cm^2 microarrays under a 14 x 14 mm glass coverslip for 6-12 hours at 60°C. Arrays are washed for 5 min at 25°C in low stringency wash buffer (1 x SSC/0.2% SDS), then for 10 min at room temperature in high stringency wash buffer (0.1 x SSC/0.2% SDS). Arrays are scanned in 0.1 x SSC using a fluorescence laser scanning device fitted with a custom filter set. Accurate differential expression measurements are obtained by taking the average of the ratios of two independent hybridizations.

Quantitative analysis of the expression of genes may also be performed with EST-related nucleic acids, fragments of EST related nucleic acids, positional segments EST-related nucleic acids, or fragments of positional segments of EST-related nucleic acids in complementary DNA arrays as described by Pietu *et al.* (*Genome Research* **6**:492-503, 1996). The EST-related nucleic acids, fragments of EST related nucleic acids, positional segments EST-related nucleic acids, or fragments of positional segments of EST-related nucleic acids thereof are PCR amplified and spotted on membranes. Then, mRNAs originating from various tissues or cells are labeled with radioactive nucleotides. After hybridization and washing in controlled conditions, the hybridized mRNAs are detected by phospho-imaging or autoradiography. Duplicate experiments are performed and a quantitative analysis of differentially expressed mRNAs is then performed.

Alternatively, expression analysis of the EST-related nucleic acids, fragments of EST related nucleic acids, positional segments EST-related nucleic acids, or fragments of positional segments of EST-related nucleic acids can be done through high density nucleotide arrays as described by Lockhart *et al.* (*Nature Biotechnology* **14**: 1675-1680, 1996) and Sosnowsky *et al.* (*Proc. Natl. Acad. Sci.* **94**:1119-1123, 1997). Oligonucleotides of 15-50 nucleotides corresponding to sequences of EST-related nucleic acids, fragments of EST related nucleic acids, positional segments EST-related nucleic acids, or fragments of positional segments of EST-related nucleic acids are synthesized directly on the chip (Lockhart *et al.*, *supra*) or synthesized and then addressed to the chip (Sosnowsky *et al.*, *supra*). Preferably, the oligonucleotides are about 20 nucleotides in length.

cDNA probes labeled with an appropriate compound, such as biotin, digoxigenin or fluorescent dye, are synthesized from the appropriate mRNA population and then

A pair of nested primers on each end is designed based on the known 5' sequence from the 5' EST or contigated consensus 5' EST and the known 3' end added by the poly dT primer used in the first strand synthesis. Software used to design primers are either based on GC content and melting temperatures of oligonucleotides, such as OSP (Illier and Green, *PCR Meth. Appl.* 1:124-128, 1991), or based on the octamer frequency disparity method (Griffais *et al.*, *Nucleic Acids Res.* 19: 3887-3891, 1991 such as PC-Rare (http://bioinformatics.weizmann.ac.il/ software/PC-Rare/doc/manuel.html).

Preferably, the nested primers at the 5' end and the nested primers at the 3' end are separated from one another by four to nine bases. These primer sequences may be selected to have melting temperatures and specificities suitable for use in PCR.

A first PCR run is performed using the outer primer from each of the nested pairs. A second PCR run is performed using the same enzyme and the inner primer from each of the nested pairs is then performed on a small sample of the first PCR product. Thereafter, the primers and remaining nucleotide monomers are removed.

2. Sequencing of Full Length Extended cDNAs or Fragments Thereof

Due to the lack of position constraints on the design of 5' nested primers compatible for PCR use using the OSP software, amplicons of two types are obtained. Preferably, the second 5' primer is located upstream of the translation initiation codon thus yielding a nested PCR product containing the entire coding sequence. Such a full length extended cDNA may be used in a direct cloning procedure. However, in some cases, the second 5' primer is located downstream of the translation initiation codon, thereby yielding a PCR product containing only part of the ORF. Such incomplete PCR products are submitted to a modified procedure described in section b below.

a) Nested PCR products containing complete ORFs

When the resulting nested PCR product contains the complete coding sequence, as predicted from the 5'EST or consensus contigated 5' EST sequence, it is cloned in an appropriate vector.

b) Nested PCR products containing incomplete ORFs

When the amplicon does not contain the complete coding sequence, intermediate steps are necessary to obtain both the complete coding sequence and a PCR product containing the full coding sequence. The complete coding sequence can

cDNA inserts, approximately 50 bp of vector DNA on each side of the cDNA insert are also sequenced.

Cloned PCR products are then entirely sequenced in order to obtain at least two sequences per clone. Preferably, the sequences are obtained from both sense and antisense strands according to the aforementioned procedure with the following modifications. First, both 5' and 3' ends of cloned PCR products are sequenced in order to confirm the identity of the clone. Second, primer walking is performed if the full coding coding region has not been obtained yet. Contigation is then performed using primer walking sequences for cloned products as well as walking sequences that have already contigated for uncloned PCR products. The sequence is considered complete when the resulting contigs include the whole coding region as well as overlapping sequences with vector DNA on both ends. All the contigated sequences for each cloned amplicon are then used to obtain a consensus sequence.

4. Selection of cloned full length sequences obtained from the 5' ESTs of the present invention

A negative selection may be performed in order to eliminate unwanted cloned sequences resulting from either contaminants or PCR artifacts as follows. Sequences matching contaminant sequences such as vector DNA, tRNA, mtRNA, rRNA sequences are discarded as well as those encoding ORF sequences exhibiting extensive homology to repeats. Sequences obtained by direct cloning using nested primers on 5' and 3' tags (section 1. case a) but lacking polyA tail may be discarded. Only ORFs containing a signal peptide and ending either before the polyA tail (case a) or before the end of the cloned 3'UTR (case b) may be selected. Then, ORFs containing unlikely mature proteins such as mature proteins which size is less than 20 amino acids or less than 25% of the immature protein size may be eliminated.

Then, for each remaining full length extended cDNA containing several ORFs, a preselection of ORFs may be performed using the following criteria. The longest ORF with a signal peptide is preferred. If the ORF sizes are similar, the chosen ORF is the one which signal peptide has the highest score according to Von Heijne method

Sequences of full length extended cDNA clones may then be compared pairwise with BLAST after masking of the repeat sequences. Sequences containing at least 90% homology over 30 nucleotides may be clustered in the same class. Each cluster may then

homologous to extended cDNAs, 5' ESTs, or consensus contigated 5' ESTs. Example 18 below provides examples of such methods.

EXAMPLE 18

Methods for Obtaining Extended cDNAs which Include the Entire Coding Region and the Authentic 5'End of the Corresponding mRNA or Nucleic Acids Homologous to Extended cDNAs, 5' ESTs or Consensus Contigated 5' ESTs

A full-length cDNA library can be made using the strategies described in Examples 1-4 above by replacing the random nonamer used in Example 2 with an oligo-dT primer. Alternatively, a cDNA library or genomic DNA library may be obtained from a commercial source or made using techniques familiar to those skilled in the art.

Such cDNA or genomic DNA libraries may be used to isolate extended cDNAs obtained from 5' ESTs or consensus contigated 5' ESTs or nucleic acids homologous to extended cDNAs, 5' ESTs, or consensus contigated 5' ESTs as follows. The cDNA library or genomic DNA library is hybridized to a detectable probe. The detectable probe may comprise at least 10, 15, 18, 20, 25, 28, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400 or 500 consecutive nucleotides of the 5' EST, consensus contigated 5' EST, or extended cDNA.

Techniques for identifying cDNA clones in a cDNA library which hybridize to a given probe sequence are disclosed in Sambrook *et al., Molecular Cloning: A Laboratory Manual 2d Ed.*, Cold Spring Harbor Laboratory Press, 1989, the disclosure of which is incorporated herein by reference. The same techniques may be used to isolate genomic DNAs.

Briefly, cDNA or genomic DNA clones which hybridize to the detectable probe are identified and isolated for further manipulation as follows. The detectable probe described in the preceding paragraph is labeled with a detectable label such as a radioisotope or a fluorescent molecule. Techniques for labeling the probe are well known and include phosphorylation with polynucleotide kinase, nick translation, *in vitro* transcription, and non radioactive techniques. The cDNAs or genomic DNAs in the library are transferred to a nitrocellulose or nylon filter and denatured. After blocking of non specific sites, the filter is incubated with the labeled probe for an amount of time

sufficient to allow binding of the probe to cDNAs or genomic DNAs containing a sequence capable of hybridizing thereto.

By varying the stringency of the hybridization conditions used to identify cDNAs or genomic DNAs which hybridize to the detectable probe, cDNAs or genomic DNAs having different levels of homology to the probe can be identified and isolated as described below.

1. Identification of cDNA or Genomic DNA Sequences Having a High Degree of Homology to the Labeled Probe

To identify cDNAs or genomic DNAs having a high degree of homology to the probe sequence, the melting temperature of the probe may be calculated using the following formulas:

For probes between 14 and 70 nucleotides in length the melting temperature (Tm) is calculated using the formula: Tm=81.5+16.6(log [Na+])+0.41(fraction G+C)-(600/N) where N is the length of the probe.

If the hybridization is carried out in a solution containing formamide, the melting temperature may be calculated using the equation Tm=81.5+16.6(log [Na+])+0.41(fraction G+C)-(0.63% formamide)-(600/N) where N is the length of the probe.

Prehybridization may be carried out in 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 100 μg denatured fragmented salmon sperm DNA or 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 100 μg denatured fragmented salmon sperm DNA, 50% formamide. The formulas for SSC and Denhardt's solutions are listed in Sambrook *et al.*, *supra.*

Hybridization is conducted by adding the detectable probe to the prehybridization solutions listed above. Where the probe comprises double stranded DNA, it is denatured before addition to the hybridization solution. The filter is contacted with the hybridization solution for a sufficient period of time to allow the probe to hybridize to extended cDNAs or genomic DNAs containing sequences complementary thereto or homologous thereto. For probes over 200 nucleotides in length, the hybridization may be carried out at 15-25°C below the Tm. For shorter probes, such as oligonucleotide probes, the hybridization may be conducted at 15-25°C below the Tm. Preferably, for hybridizations in 6X SSC, the hybridization is conducted at approximately 68°C. Preferably, for hybridizations in 50% formamide containing solutions, the hybridization is conducted at approximately 42°C.

All of the foregoing hybridizations would be considered to be under "stringent" conditions.

Following hybridization, the filter is washed in 2X SSC, 0.1% SDS at room temperature for 15 minutes. The filter is then washed with 0.1X SSC, 0.5% SDS at room temperature for 30 minutes to 1 hour. Thereafter, the solution is washed at the hybridization temperature in 0.1X SSC, 0.5% SDS. A final wash is conducted in 0.1X SSC at room temperature.

cDNAs or genomic DNAs which have hybridized to the probe are identified by autoradiography or other conventional techniques.

2. Obtaining cDNA or Genomic DNA Sequences Having Lower Degrees of Homology to the Labeled Probe

The above procedure may be modified to identify cDNAs or genomic DNAs having decreasing levels of homology to the probe sequence. For example, to obtain cDNAs or genomic DNAs of decreasing homology to the detectable probe, less stringent conditions may be used. For example, the hybridization temperature may be decreased in increments of 5°C from 68°C to 42°C in a hybridization buffer having a sodium concentration of approximately 1M. Following hybridization, the filter may be washed with 2X SSC, 0.5% SDS at the temperature of hybridization. These conditions are considered to be "moderate" conditions above 50°C and "low" conditions below 50°C.

Alternatively, the hybridization may be carried out in buffers, such as 6X SSC, containing formamide at a temperature of 42°C. In this case, the concentration of formamide in the hybridization buffer may be reduced in 5% increments from 50% to 0% to identify clones having decreasing levels of homology to the probe. Following hybridization, the filter may be washed with 6X SSC, 0.5% SDS at 50°C. These conditions are considered to be "moderate" conditions above 25% formamide and "low" conditions below 25% formamide.

cDNAs or genomic DNAs which have hybridized to the probe are identified by autoradiography.

3. Determination of the Degree of Homology between the Obtained cDNAs or Genomic DNAs and 5'ESTs, Consensus Contigated 5'ESTs, or Extended cDNAs or Between the

the statistical significance formula of Karlin (see, *e.g.*, Karlin and Altschul, 1990, *Proc. Natl. Acad. Sci. USA 87*:2267-2268).

The parameters used with the above algorithms may be adapted depending on the sequence length and degree of homology studied. In some embodiments, the parameters may be the default parameters used by the algorithms in the absence of instructions from the user.

In some embodiments, the level of homology between the hybridized nucleic acid and the extended cDNA, 5'EST, or 5' consensus contigated EST from which the probe was derived may be determined using the FASTDB algorithm described in Brutlag et al. Comp. App. Biosci. 6:237-245, 1990. In such analyses the parameters may be selected as follows: Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the length of the sequence which hybridizes to the probe, whichever is shorter. Because the FASTDB program does not consider 5′ or 3′ truncations when calculating homology levels, if the sequence which hybridizes to the probe is truncated relative to the sequence of the extended cDNA, 5'EST, or consensus contigated 5'EST from which the probe was derived the homology level is manually adjusted by calculating the number of nucleotides of the extended cDNA, 5'EST, or consensus contigated 5' EST which are not matched or aligned with the hybridizing sequence, determining the percentage of total nucleotides of the hybridizing sequence which the non-matched or non-aligned nucleotides represent, and subtracting this percentage from the homology level. For example, if the hybridizing sequence is 700 nucleotides in length and the extended cDNA, 5'EST, or consensus contigated 5' EST sequence is 1000 nucleotides in length wherein the first 300 bases at the 5′ end of the extended cDNA, 5'EST, or consensus contigated 5' EST are absent from the hybridizing sequence, and wherein the overlapping 700 nucleotides are identical, the homology level would be adjusted as follows. The non-matched, non-aligned 300 bases represent 30% of the length of the extended cDNA, 5'EST, or consensus contigated 5' EST. If the overlapping 700 nucleotides are 100% identical, the adjusted homology level would be 100-30=70% homology. It should be noted that the preceding adjustments are only made when the non-matched or non-aligned nucleotides are at the 5′ or 3′ ends. No adjustments are made if the non-matched or non-aligned sequences are internal or under any other conditions.

For example, using the above methods, nucleic acids having at least 95% nucleic acid homology, at least 96% nucleic acid homology, at least 97% nucleic acid homology, at least 98% nucleic acid homology, at least 99% nucleic acid homology, or more than 99% nucleic acid homology to the extended cDNA, 5'EST, or consensus contigated 5' EST from which the probe was derived may be obtained and identified. Such nucleic acids may be allelic variants or related nucleic acids from other species. Similarly, by using progressively less stringent hybridization conditions one can obtain and identify nucleic acids having at least 90%, at least 85%, at least 80% or at least 75% homology to the extended cDNA, 5'EST, or consensus contigated 5' EST from which the probe was derived.

Using the above methods and algorithms such as FASTA with parameters depending on the sequence length and degree of homology studied, for example the default parameters used by the algorithms in the absence of instructions from the user, one can obtain nucleic acids encoding proteins having at least 99%, at least 98%, at least 97%, at least 96%, at least 95%, at least 90%, at least 85%, at least 80% or at least 75% homology to the protein encoded by the extended cDNA, 5'EST, or consensus contigated 5' EST from which the probe was derived. In some embodiments, the homology levels can be determined using the "default" opening penalty and the "default" gap penalty, and a scoring matrix such as PAM 250 (a standard scoring matrix; see Dayhoff et al., in: Atlas of Protein Sequence and Structure, Vol. 5, Supp. 3 (1978)).

Alternatively, the level of polypeptide homology may be determined using the FASTDB algorithm described by Brutlag et al. Comp. App. Biosci. 6:237-245, 1990. In such analyses the parameters may be selected as follows: Matrix=PAM 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=Sequence Length, Gap Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the length of the homologous sequence, whichever is shorter. If the homologous amino acid sequence is shorter than the amino acid sequence encoded by the extended cDNA, 5'EST, or consensus contigated 5' EST as a result of an N terminal and/or C terminal deletion the results may be manually corrected as follows. First, the number of amino acid residues of the amino acid sequence encoded by the extended cDNA, 5'EST, or consensus contigated 5' EST which are not matched or aligned with the homologous sequence is determined. Then, the percentage of the length of the sequence

Furthermore, the polypeptides encoded by the extended or full-length cDNAs may be screened for the presence of known structural or functional motifs or for the presence of signatures, small amino acid sequences which are well conserved amongst the members of a protein family. The results obtained for the polypeptides encoded by a few full-length cDNAs derived from 5'ESTs that were screened for the presence of known protein signatures and motifs using the Proscan software from the GCG package and the Prosite 15.0 database are provided below.

The protein of SEQ ID NO: 8 encoded by the full-length cDNA SEQ ID NO: 7 (internal designation 78-8-3-E6-CL0_1C) and expressed in adult prostate belong to the phosphatidylethanolamine-binding protein from which it exhibits the characteristic PROSITE signature from positions 90 to 112. Proteins from this widespread family, from nematodes to fly, yeast, rodent and primate species, bind hydrophobic ligands such as phospholipids and nucleotides. They are mostly expressed in brain and in testis and are thought to play a role in cell growth and/or maturation, in regulation of the sperm maturation, motility and in membrane remodeling. They may act either through signal transduction or through oxidoreduction reactions (for a review see Schoentgen and Jollès, *FEBS Letters*, **369** :22-26 (1995)). Taken together, these data suggest that the protein of SEQ ID NO: 8 may play a role in cell growth, maturation and in membrane remodeling and/or may be related to male fertility. Thus, these protein may be useful in diagnosing and/or treating cancer, neurodegenerative diseases, and/or disorders related to male fertility and sterility.

The protein of SEQ ID NO :10 encoded by the full-length cDNA SEQ ID NO:9 (internal designation 108-013-5-O-H9-FLC) shows homologies with a family of lysophospholipases conserved among eukaryotes (yeast, rabbit, rodents and human). In addition, some members of this family exhibit a calcium-independent phospholipase A2 activity (Portilla *et al, J. Am. Soc. Nephro.*, **9** :1178-1186 (1998)). All members of this family exhibit the active site consensus GXSXG motif of carboxylesterases that is also found in the protein of SEQ ID NO :10 (position 54 to 58). In addition, this protein may be a membrane protein with one transmembrane domain as predicted by the software TopPred II (Claros and von Heijne, *CABIOS applic. Notes*, **10** :685-686 (1994)). Taken together, these data suggest that the protein of SEQ ID NO:10 may play a role in fatty acid metabolism, probably as a phospholipase. Thus, this protein or part

therein, may be useful in diagnosing and/or treating several disorders including, but not limited to, cancer, diabetes, and neurodegenerative disorders such as Parkinson's and Alzheimer's diseases. It may also be useful in modulating inflammatory responses to infectious agents and/or to suppress graft rejection.

The protein of SEQ ID NO: 12 encoded by the full-length cDNA SEQ ID NO: 11 (internal designation 108-004-5-0-D10-FLC) shows remote homology to a subfamily of beta4-galactosyltransferases widely conserved in animals (human, rodents, cow and chicken). Such enzymes, usually type II membrane proteins located in the endoplasmic reticulum or in the Golgi apparatus, catalyzes the biosynthesis of glycoproteins, glycolipid glycans and lactose. Their characteristic features defined as those of subfamily A in Breton *et al, J. Biochem.*, **123**:1000-1009 (1998) are pretty well conserved in the protein of SEQ ID NO: 12, especially the region I containing the DVD motif (positions 163-165) thought to be involved either in UDP binding or in the catalytic process itself. In addition, the protein of SEQ ID NO: 12 has the typical structure of a type II protein. Indeed, it contains a short 28-amino-acid-long N-terminal tail, a transmembrane segment from positions 29 to 49 and a large 278-amino-acid-long C-terminal tail as predicted by the software TopPred II (Claros and von Heijne, *CABIOS applic. Notes*, **10** :685-686 (1994)). Taken together, these data suggest that the protein of SEQ ID NO: 12 may play a role in the biosynthesis of polysaccharides, and of the carbohydrate moieties of glycoproteins and glycolipids and/or in cell-cell recognition. Thus, this protein may be useful in diagnosing and/or treating several types of disorders including, but not limited to, cancer, atherosclerosis, cardiovascular disorders, autoimmune disorders and rheumatic diseases including rheumatoid arthritis.

The protein of SEQ ID NO: 14 encoded by the full-length cDNA SEQ ID NO: 13 (internal designation 108-009-5-0-A2-FLC) shows extensive homology to the bZIP family of transcription factors, and especially to the human luman protein (Lu *et al.*, *Mol. Cell. Biol.*, **17** :5117-5126 (1997))). The match include the whole bZIP domain composed of a basic DNA-binding domain and of a leucine zipper allowing protein dimerization. The basic domain is conserved in the protein of SEQ ID NO: 14 as shown by the characteristic PROSITE signature (positions 224-237) except for a conservative substitution of a glutamic acid with an aspartic acid in position 233. The typical PROSITE signature for leucine zipper is also present (positions 259 to 280).

Following the incubation, the cells are washed to remove non-specifically bound proteins or polypeptides. The specifically bound labeled proteins or polypeptides are detected by autoradiography. Alternatively, unlabeled proteins or polypeptides may be incubated with the cells and detected with antibodies having a detectable label, such as a fluorescent molecule, attached thereto.

Specificity of cell surface binding may be analyzed by conducting a competition analysis in which various amounts of unlabeled protein or polypeptide are incubated along with the labeled protein or polypeptide. The amount of labeled protein or polypeptide bound to the cell surface decreases as the amount of competitive unlabeled protein or polypeptide increases. As a control, various amounts of an unlabeled protein or polypeptide unrelated to the labeled protein or polypeptide is included in some binding reactions. The amount of labeled protein or polypeptide bound to the cell surface does not decrease in binding reactions containing increasing amounts of unrelated unlabeled protein, indicating that the protein or polypeptide encoded by the nucleic acid binds specifically to the cell surface.

As discussed above, human proteins have been shown to have a number of important physiological effects and, consequently, represent a valuable therapeutic resource. The human proteins or polypeptides made as described above may be evaluated to determine their physiological activities as described below.

EXAMPLE 22

Assaying the Expressed Proteins or Polypeptides for Cytokine, Cell Proliferation or Cell Differentiation Activity

As discussed above, some human proteins act as cytokines or may affect cellular proliferation or differentiation. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of a protein or polypeptide of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M^+ (preB M^+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7c and CMK. The proteins or polypeptides prepared as described above may be evaluated for their ability to regulate T

proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases caused by viral, bacterial, fungal or other infection may be treatable using the protein or polypeptide including infections by HIV, hepatitis viruses, herpesviruses, mycobacteria, Leishmania spp., plamodium. and various fungal infections such as candidiasis. Of course, in this regard, a protein or polypeptide may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

Alternatively, the proteins or polypeptides prepared as described above may be used in treatment of autoimmune disorders including, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein or polypeptide may also to be useful in the treatment of allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using the protein or polypeptide.

Using the proteins or polypeptides of the invention it may also be possible to regulate immune responses either up or down. Down regulation may involve inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response. The functions of activated T-cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active non-antigen-specific process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after the end of exposure to the tolerizing agent. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions, such as, for example, B7 costimulation), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a molecule which inhibits or blocks interaction of a B7 lymphocyte antigen with its natural ligand(s) on immune cells (such as a soluble, monomeric form of a peptide having B7-2 activity alone or in conjunction with a monomeric form of a peptide having an activity of another B lymphocyte antigen (e.g., B7-1, B7-3) or blocking antibody), prior to transplantation, can lead to the binding of the molecule to the natural ligand(s) on the immune cells without transmitting the corresponding costimulatory signal. Blocking B lymphocyte antigen function in this matter prevents cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, the lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular blocking reagents in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins *in vivo* as described in Lenschow *et al.*, *Science* **257**:789-792 (1992) and Turka *et al.*, *Proc. Natl. Acad. Sci USA*, **89**:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed., *Fundamental Immunology*, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of blocking B lymphocyte antigen function *in vivo* on the development of that disease.

In another application, upregulation or enhancement of antigen function (preferably B lymphocyte antigen function) may be useful in the induction of tumor immunity. Tumor cells (e.g., sarcoma, melanoma, lymphoma, leukemia, neuroblastoma, carcinoma) transfected with one of the above-described nucleic acids encoding a protein or polypeptide can be administered to a subject to overcome tumor-specific tolerance in the subject. If desired, the tumor cell can be transfected to express a combination of peptides. For example, tumor cells obtained from a patient can be transfected ex vivo with an expression vector directing the expression of a peptide having B7-2-like activity alone, or in conjunction with a peptide having B7-1-like activity and/or B7-3-like activity. The transfected tumor cells are returned to the patient to result in expression of the peptides on the surface of the transfected cell. Alternatively, gene therapy techniques can be used to target a tumor cell for transfection *in vivo*.

The presence of the protein or polypeptide encoded by the nucleic acids described above having the activity of a B lymphocyte antigen(s) on the surface of the tumor cell provides the necessary costimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack or which fail to reexpress sufficient amounts of MHC class I or MHC class II molecules can be transfected with nucleic acids encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I α chain and β_2 microglobulin or an MHC class II α chain and an MHC class II β chain to thereby express MHC class I or MHC class II proteins on the cell surface, respectively. Expression of the appropriate MHC class I or class II molecules in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a nucleic acid encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a protein or polypeptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject. Alternatively, as described in more detail below, nucleic acids encoding these immune system regulator proteins or polypeptides or nucleic acids regulating the expression of such

proteins or polypeptides may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

EXAMPLE 24

Assaying the Expressed Proteins or Polypeptides for Hematopoiesis Regulating Activity

The proteins or polypeptides encoded by the nucleic acids described above may also be evaluated for their hematopoiesis regulating activity. For example, the effect of the proteins or polypeptides on embryonic stem cell differentiation may be evaluated. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Johansson *et al. Cell. Biol.* **15**:141-151, 1995; Keller *et al.*, *Mol. Cell. Biol.* **13**:473-486, 1993; McClanahan *et al.*, *Blood* **81**:2903-2915, 1993.

The proteins or polypeptides encoded by the nucleic acids described above may also be evaluated for their influence on the lifetime of stem cells and stem cell differentiation. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Freshney, M.G. Methylcellulose Colony Forming Assays, in Culture of Hematopoietic Cells. R.I. Freshney, et al. Eds. pp. 265-268, Wiley-Liss, Inc., New York, NY. 1994; Hirayama et al., *Proc. Natl. Acad. Sci. USA* 89:5907-5911, 1992; McNiece, I.K. and Briddell, R.A. Primitive Hematopoietic Colony Forming Cells with High Proliferative Potential, in Culture of Hematopoietic Cells. R.I. Freshney, et al. eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, NY. 1994; Neben et al., *Experimental Hematology* 22:353-359, 1994; Ploemacher, R.E. Cobblestone Area Forming Cell Assay, In Culture of Hematopoietic Cells. R.I. Freshney, et al. Eds. pp. 1-21, Wiley-Liss, Inc., New York, NY. 1994; Spooncer, E., Dexter, M. and Allen, T. Long Term Bone Marrow Cultures in the Presence of Stromal Cells, in Culture of Hematopoietic Cells. R.I. Freshney, et al. Eds. pp. 163-179, Wiley-Liss, Inc., New York, NY. 1994; and Sutherland, H.J. Long Term Culture Initiating Cell Assay, in Culture of Hematopoietic Cells. R.I. Freshney, et al. Eds. pp. 139-162, Wiley-Liss, Inc., New York, NY. 1994.

Those proteins or polypeptides which exhibit hematopoiesis regulatory activity may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of hematopoeisis is beneficial. For example, a protein or polypeptide of the

induces tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a protein or polypeptide of the present invention contributes to the repair of tendon or ligaments defects of congenital, traumatic or other origin and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The proteins or polypeptides of the present invention may provide an environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors ex vivo for return *in vivo* to effect tissue repair. The proteins or polypeptides of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The therapeutic compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The proteins or polypeptides of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e., for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a protein or polypeptide may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a protein or polypeptide of the invention.

Taub *et al. J. Clin. Invest.* **95**:1370-1376, 1995; Lind *et al. APMIS* **103**:140-146, 1995; Muller *et al. Eur. J. Immunol.* **25**:1744-1748; Gruber *et al. J. Immunol.* **152**:5860-5867, 1994; Johnston *et al., J Immunol.* **153**:1762-1768, 1994.

Those proteins or polypeptides which exhibit activity as reproductive hormones or regulators of cell movement may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of reproductive hormones are beneficial. For example, a protein or polypeptide may exhibit activin- or inhibin-related activities. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins are characterized by their ability to stimulate the release of FSH. Thus, a protein or polypeptide of the present invention, alone or in heterodimers with a member of the inhibin α family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the protein or polypeptide of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin-B group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, United States Patent 4,798,885, the disclosure of which is incorporated herein by reference. A protein or polypeptide of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as cows, sheep and pigs.

Alternatively, as described in more detail below, nucleic acids encoding reproductive hormone regulating activity proteins or polypeptides or nucleic acids regulating the expression of such proteins or polypeptides may be introduced into appropriate host cells to increase or decrease the expression of the proteins or polypeptides as desired.

EXAMPLE 27

Assaying the Expressed Proteins or Polypeptides For Chemotactic/Chemokinetic Activity

The proteins or polypeptides of the present invention may also be evaluated for chemotactic/chemokinetic activity. For example, a protein or polypeptide of the present invention may have chemotactic or chemokinetic activity (e.g., act as a chemokine) for

metabolism, catabolism, anabolism, processing, utilization, storage or climination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein. Alternatively, as described in more detail below, nucleic acids encoding proteins or polypeptides involved in any of the above mentioned activities or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins or polypeptides as desired.

EXAMPLE 32

Identification of Proteins or Polypeptides which Interact with Proteins or Polypeptides of the Present Invention

Proteins or polypeptides which interact with the proteins or polypeptides of the present invention, such as receptor proteins, may be identified using two hybrid systems such as the Matchmaker Two Hybrid System 2 (Catalog No. K1604-1, Clontech). As described in the manual accompanying the kit which is incorporated herein by reference, nucleic acids encoding the proteins or polypeptides of the present invention, are inserted into an expression vector such that they are in frame with DNA encoding the DNA binding domain of the yeast transcriptional activator GAL4. cDNAs in a cDNA library which encode proteins or polypeptides which might interact with the proteins or polypeptides of the present invention are inserted into a second expression vector such that they are in frame with DNA encoding the activation domain of GAL4. The two expression plasmids are transformed into yeast and the yeast are plated on selection medium which selects for expression of selectable markers on each of the expression vectors as well as GAL4

In other embodiments, the antibodies may be capable of specifically binding to an EST-related polypeptide, fragment of an EST-related polypeptide, positional segment of an EST-related polypeptide or fragment of a positional segment of an EST-related polypeptide. In some embodiments, the antibody may be capable of binding an antigenic determinant or an epitope in an EST-related polypeptide, fragment of an EST-related polypeptide, positional segment of an EST-related polypeptide or fragment of a positional segment of an EST-related polypeptide.

In the case of secreted proteins, the antibodies may be capable of binding a full-length protein encoded by a nucleic acid of the present invention, a mature protein (i.e. the protein generated by cleavage of the signal peptide) encoded by a nucleic acid of the present invention, or a signal peptide encoded by a nucleic acid of the present invention.

EXAMPLE 33

Production of an Antibody to a Human Polypeptide or Protein

The above described EST-related nucleic acids, fragments of EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids or nucleic acids encoding EST-related polypeptides, fragments of EST-related polypeptides, positional segments of EST-related polypeptides or fragments of positional segments of EST-related polypeptides are operably linked to promoters and introduced into cells as described above.

In the case of secreted proteins, nucleic acids encoding the full protein (i.e. the mature protein and the signal peptide), nucleic acids encoding the mature protein (i.e. the protein generated by cleavage of the signal peptide), or nucleic acids encoding the signal peptide are operably linked to promoters and introduced into cells as described above.

The encoded proteins or polypeptides are then substantially purified or isolated as described above. The concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few μg/ml. Monoclonal or polyclonal antibody to the protein or polypeptide can then be prepared as follows:

1. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes of any of the proteins or polypeptides identified and isolated as described can be prepared from murine hybridomas according to the

classical method of Kohler, and Milstein, *Nature* **256**:495 (1975) or derivative methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein or peptides derived therefrom over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as Elisa, as originally described by Engvall, *Meth. Enzymol.* **70**:419 (1980), the disclosure of which is incorporated herein by reference and derivative methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. *et al.* in *Basic Methods in Molecular Biology* Elsevier, New York. Section 21-2, the disclosure of which is incorporated herein by reference.

2. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogenous epitopes of a single protein or polypeptide can be prepared by immunizing suitable animals with the expressed protein or peptides derived therefrom, which can be unmodified or modified to enhance immunogenicity. Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than others and may require the use of carriers and adjuvant. Also, host animals response vary depending on site of inoculations and doses, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis. *et al.J. Clin. Endocrinol. Metab.* **33**:988-991 (1971) , the disclosure of which is incorporated herein by reference.

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, *et al.*, Chap. 19 in: *Handbook of Experimental Immunology* D.

A panel of probes based on the sequences of the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids are radioactively or colorimetrically labeled using methods known in the art, such as nick translation or end labeling, and hybridized to the Southern blot using techniques known in the art (Davis *et al.*, supra). Preferably, the probe is at least 10, 12, 15, 18, 20, 25, 28, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400 or 500 nucleotides in length. Preferably, the probes are at least 10, 12, 15, 18, 20, 25, 28, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400 or 500 nucleotides in length. In some embodiments, the probes are oligonucleotides which are 40 nucleotides in length or less.

Preferably, at least 5 to 10 of these labeled probes are used, and more preferably at least about 20 or 30 are used to provide a unique pattern. The resultant bands appearing from the hybridization of a large sample of EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids will be a unique identifier. Since the restriction enzyme cleavage will be different for every individual, the band pattern on the Southern blot will also be unique. Increasing the number of probes will provide a statistically higher level of confidence in the identification since there will be an increased number of sets of bands used for identification.

EXAMPLE 39

Dot Blot Identification Procedure

Another technique for identifying individuals using the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids disclosed herein utilizes a dot blot hybridization technique.

Genomic DNA is isolated from nuclei of subject to be identified. Probes are prepared that correspond to at least 10, preferably 50 sequences from the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids. The probes are used to hybridize to the genomic DNA through conditions known to those in the art. The oligonucleotides are end labeled with P^{32} using polynucleotide kinase (Pharmacia). Dot Blots are created by spotting the genomic DNA onto nitrocellulose or the like using a vacuum dot blot manifold (BioRad, Richmond California). The nitrocellulose filter containing the genomic sequences is baked

DNA are loaded into wells and separated on 0.8% agarose gels. The gels are transferred onto nitrocellulose using standard Southern blotting techniques.

10 ng of each of the oligonucleotides are pooled and end-labeled with P^{32}. The nitrocellulose is prehybridized with blocking solution and hybridized with the labeled probes. Following hybridization and washing, the nitrocellulose filter is exposed to X-Omat AR X-ray film. The resulting hybridization pattern will be unique for each individual.

It is additionally contemplated within this example that the number of probe sequences used can be varied for additional accuracy or clarity.

In addition to their applications in forensics and identification, EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids may be mapped to their chromosomal locations. Example 41 below describes radiation hybrid (RH) mapping of human chromosomal regions using EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids. Example 42 below describes a representative procedure for mapping EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids to their locations on human chromosomes. Example 43 below describes mapping of EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids on metaphase chromosomes by Fluorescence In Situ Hybridization (FISH).

2. Use of EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids in Chromosome Mapping

EXAMPLE 41

Radiation hybrid mapping of EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids to the human genome

Radiation hybrid (RH) mapping is a somatic cell genetic approach that can be used for high resolution mapping of the human genome. In this approach, cell lines containing one or more human chromosomes are lethally irradiated, breaking each chromosome into

fragments whose size depends on the radiation dose. These fragments are rescued by fusion with cultured rodent cells, yielding subclones containing different portions of the human genome. This technique is described by Benham *et al.* (*Genomics* **4**:509-517, 1989) and Cox *et al.*, (*Science* **250**:245-250, 1990), the entire contents of which are hereby incorporated by reference. The random and independent nature of the subclones permits efficient mapping of any human genome marker. Human DNA isolated from a panel of 80-100 cell lines provides a mapping reagent for ordering EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids. In this approach, the frequency of breakage between markers is used to measure distance, allowing construction of fine resolution maps as has been done using conventional ESTs (Schuler *et al.*, *Science* **274**:540-546, 1996, hereby incorporated by reference).

RH mapping has been used to generate a high-resolution whole genome radiation hybrid map of human chromosome 17q22-q25.3 across the genes for growth hormone (GH) and thymidine kinase (TK) (Foster *et al.*, *Genomics* **33**:185-192, 1996), the region surrounding the Gorlin syndrome gene (Obermayr *et al.*, *Eur. J. Hum. Genet.* **4**:242-245, 1996), 60 loci covering the entire short arm of chromosome 12 (Raeymaekers *et al.*, *Genomics* **29**:170-178, 1995), the region of human chromosome 22 containing the neurofibromatosis type 2 locus (Frazer *et al.*, *Genomics* **14**:574-584, 1992) and 13 loci on the long arm of chromosome 5 (Warrington *et al.*, *Genomics* **11**:701-708, 1991).

EXAMPLE 42

Mapping of EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids to Human Chromosomes using PCR techniques

EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids may be assigned to human chromosomes using PCR based methodologies. In such approaches, oligonucleotide primer pairs are designed from EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids to minimize the chance of amplifying through an intron. Preferably, the oligonucleotide primers are 18-23 bp in length and are designed for PCR amplification. The creation of

Slides kept at -20°C are treated for 1 h at 37°C with RNase A (100 μg/ml), rinsed three times in 2 X SSC and dehydrated in an ethanol series. Chromosome preparations are denatured in 70% formamide, 2 X SSC for 2 min at 70°C, then dehydrated at 4°C. The slides are treated with proteinase K (10 μg/100 ml in 20 mM Tris-HCl, 2 mM $CaCl_2$) at 37°C for 8 min and dehydrated. The hybridization mixture containing the probe is placed on the slide, covered with a coverslip, sealed with rubber cement and incubated overnight in a humid chamber at 37°C. After hybridization and post-hybridization washes, the biotinylated probe is detected by avidin-FITC and amplified with additional layers of biotinylated goat anti-avidin and avidin-FITC. For chromosomal localization, fluorescent R-bands are obtained as previously described (Cherif *et al.*, *supra.*). The slides are observed under a LEICA fluorescence microscope (DMRXA). Chromosomes are counterstained with propidium iodide and the fluorescent signal of the probe appears as two symmetrical yellow-green spots on both chromatids of the fluorescent R-band chromosome (red). Thus, a particular EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids may be localized to a particular cytogenetic R-band on a given chromosome.

Once the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids have been assigned to particular chromosomes using the techniques described in Examples 41-43 above, they may be utilized to construct a high resolution map of the chromosomes on which they are located or to identify the chromosomes in a sample.

EXAMPLE 44

Use of EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids to Construct or Expand Chromosome Maps

Chromosome mapping involves assigning a given unique sequence to a particular chromosome as described above. Once the unique sequence has been mapped to a given chromosome, it is ordered relative to other unique sequences located on the same chromosome. One approach to chromosome mapping utilizes a series of yeast artificial chromosomes (YACs) bearing several thousand long inserts derived from the chromosomes of the organism from which the EST-related nucleic acids, positional

segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids are obtained. This approach is described in Ramaiah Nagaraja *et al., Genome Research* 7:210-222, March 1997, the disclosure of which is incorporated herein by reference. Briefly, in this approach each chromosome is broken into overlapping pieces which are inserted into the YAC vector. The YAC inserts are screened using PCR or other methods to determine whether they include the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids whose position is to be determined. Once an insert has been found which includes the 5'EST, the insert can be analyzed by PCR or other methods to determine whether the insert also contains other sequences known to be on the chromosome or in the region from which the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids was derived. This process can be repeated for each insert in the YAC library to determine the location of each of the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids relative to one another and to other known chromosomal markers. In this way, a high resolution map of the distribution of numerous unique markers along each of the organisms chromosomes may be obtained.

As described in Example 45 below EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids may also be used to identify genes associated with a particular phenotype, such as hereditary disease or drug response.

3. Use of EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids Gene Identification

EXAMPLE 45

Identification of genes associated with hereditary diseases or drug response

This example illustrates an approach useful for the association of EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids with particular phenotypic characteristics. In this example, a particular EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids is used as a

The secretion vector may be DNA or RNA and may integrate into the chromosome of the host, be stably maintained as an extrachromosomal replicon in the host, be an artificial chromosome, or be transiently present in the host. Preferably, the secretion vector is maintained in multiple copies in each host cell. As used herein, multiple copies means at least 2, 5, 10, 20, 25, 50 or more than 50 copies per cell. In some embodiments, the multiple copies are maintained extrachromosomally. In other embodiments, the multiple copies result from amplification of a chromosomal sequence.

Many nucleic acid backbones suitable for use as secretion vectors are known to those skilled in the art, including retroviral vectors, SV40 vectors, Bovine Papilloma Virus vectors, yeast integrating plasmids, yeast episomal plasmids, yeast artificial chromosomes, human artificial chromosomes, P element vectors, baculovirus vectors, or bacterial plasmids capable of being transiently introduced into the host.

The secretion vector may also contain a polyA signal such that the polyA signal is located downstream of the gene inserted into the secretion vector.

After the gene encoding the protein for which secretion is desired is inserted into the secretion vector, the secretion vector is introduced into the host cell, tissue, or organism using calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection, viral particles or as naked DNA. The protein encoded by the inserted gene is then purified or enriched from the supernatant using conventional techniques such as ammonium sulfate precipitation, immunoprecipitation, immunoaffinitychromatography, size exclusion chromatography, ion exchange chromatography, and HPLC. Alternatively, the secreted protein may be in a sufficiently enriched or pure state in the supernatant or growth media of the host to permit it to be used for its intended purpose without further enrichment.

The signal sequences may also be inserted into vectors designed for gene therapy. In such vectors, the signal sequence is operably linked to a promoter such that mRNA transcribed from the promoter encodes the signal peptide. A cloning site is located downstream of the signal sequence such that a gene encoding a protein whose secretion is desired may readily be inserted into the vector and fused to the signal sequence. The vector is introduced into an appropriate host cell. The protein expressed from the promoter is secreted extracellularly, thereby producing a therapeutic effect.

EXAMPLE 47

Fusion Vectors

The EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids may be used to construct fusion vectors for the expression of chimeric polypeptides. The chimeric polypeptides comprise a first polypeptide portion and a second polypeptide portion. In the fusion vectors of the present invention, nucleic acids encoding the first polypeptide portion and the second polypeptide portion are joined in frame with one another so as to generate a nucleic acid encoding the chimeric polypeptide. The nucleic acid encoding the chimeric polypeptide is operably linked to a promoter which directs the expression of an mRNA encoding the chimeric polypeptide. The promoter may be in any of the expression vectors described herein including those described in Examples 20 and 46.

Preferably, the fusion vector is maintained in multiple copies in each host cell. In some embodiments, the multiple copies are maintained extrachromosomally. In other embodiments, the multiple copies result from amplification of a chromosomal sequence.

The first polypeptide portion may comprise any of the polypeptides encoded by the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids. In some embodiments, the first polypeptide portion may be one of the EST-related polypeptides, fragments of EST-related polypeptides, positional segments of EST-related polypeptides, or fragments of positional segments of EST-related polypeptides.

The second polypeptide portion may comprise any polypeptide of interest. In some embodiments, the second polypeptide portion may comprise a polypeptide having a detectable enzymatic activity such as green fluorescent protein or β galactosidase. Chimeric polypeptides in which the second polypeptide portion comprises a detectable polypeptide may be used to determine the intracellular localization of the first polypeptide portion. In such procedures, the fusion vector encoding the chimeric polypeptide is introduced into a host cell under conditions which facilitate the expression of the chimeric polypeptide. Where appropriate, the cells are treated with a detection reagent which is visible under the microscope following a catalytic reaction with the detectable polypeptide and the cellular location of the detection reagent is determined. For example, if the polypeptide having a detectable enzymatic activity is β galactosidase, the cells may be

treated with Xgal. Alternatively, where the detectable polypeptide is directly detectable without the addition of a detection reagent, the intracellular location of the chimeric polypeptide is determined by performing microscopy under conditions in which the dectable polypeptide is visible. For example, if the detectable polypeptide is green fluorescent protein or a modified version thereof, microscopy is performed by exposing the host cells to light having an appropriate wavelength to cause the green fluorescent protein or modified version thereof to fluoresce.

Alternatively, the second polypeptide portion may comprise a polypeptide whose isolation, purification, or enrichment is desired. In such embodiments, the isolation, purification, or enrichment of the second polypeptide portion may be achieved by performing the immunoaffinity chromatography procedures described below using an immunoaffinity column having an antibody directed against the first polypeptide portion coupled thereto.

The proteins encoded by the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids or the EST-related polypeptides, fragments of EST-related polypeptides, positional segments of EST-related polypeptides, or fragments of positional segments of EST-related polypeptides may also be used to generate antibodies as explained in Examples 20 and 33 in order to identify the tissue type or cell species from which a sample is derived as described in Example 48.

EXAMPLE 48

Identification of Tissue Types or Cell Species by Means of Labeled Tissue Specific Antibodies

Identification of specific tissues is accomplished by the visualization of tissue specific antigens by means of antibody preparations according to Examples 20 and 33 which are conjugated, directly or indirectly to a detectable marker. Selected labeled antibody species bind to their specific antigen binding partner in tissue sections, cell suspensions, or in extracts of soluble proteins from a tissue sample to provide a pattern for qualitative or semi-qualitative interpretation.

Antisera for these procedures must have a potency exceeding that of the native preparation, and for that reason, antibodies are concentrated to a mg/ml level by isolation

of the gamma globulin fraction, for example, by ion-exchange chromatography or by ammonium sulfate fractionation. Also, to provide the most specific antisera, unwanted antibodies, for example to common proteins, must be removed from the gamma globulin fraction, for example by means of insoluble immunoabsorbents, before the antibodies are labeled with the marker. Either monoclonal or heterologous antisera is suitable for either procedure.

1. Immunohistochemical Techniques

Purified, high-titer antibodies, prepared as described above, are conjugated to a detectable marker, as described, for example, by Fudenberg, H., Chap. 26 in: *Basic 503 Clinical Immunology*, 3rd Ed. Lange, Los Altos, California (1980) or Rose,. *et al.*, Chap. **12** in: *Methods in Immunodiagnosis*, 2d Ed. John Wiley and Sons, New York (1980), the disclosures of which are incorporated herein by reference.

A fluorescent marker, either fluorescein or rhodamine, is preferred, but antibodies can also be labeled with an enzyme that supports a color producing reaction with a substrate, such as horseradish peroxidase. Markers can be added to tissue-bound antibody in a second step, as described below. Alternatively, the specific antitissue antibodies can be labeled with ferritin or other electron dense particles, and localization of the ferritin coupled antigen-antibody complexes achieved by means of an electron microscope. In yet another approach, the antibodies are radiolabeled, with, for example ^{125}I, and detected by overlaying the antibody treated preparation with photographic emulsion.

Preparations to carry out the procedures can comprise monoclonal or polyclonal antibodies to a single protein or peptide identified as specific to a tissue type, for example, brain tissue, or antibody preparations to several antigenically distinct tissue specific antigens can be used in panels, independently or in mixtures, as required.

Tissue sections and cell suspensions are prepared for immunohistochemical examination according to common histological techniques. Multiple cryostat sections (about 4 μm, unfixed) of the unknown tissue and known control, are mounted and each slide covered with different dilutions of the antibody preparation. Sections of known and unknown tissues should also be treated with preparations to provide a positive control, a negative control, for example, pre-immune sera, and a control for non-specific staining, for example, buffer.

EXAMPLE 50

Immunoaffinity Chromatography

Antibodies prepared as described above are coupled to a support. Preferably, the antibodies are monoclonal antibodies, but polyclonal antibodies may also be used. The support may be any of those typically employed in immunoaffinity chromatography, including Sepharose CL-4B (Pharmacia, Piscataway, NJ), Sepharose CL-2B (Pharmacia, Piscataway, NJ), Affi-gel 10 (Biorad, Richmond, CA), or glass beads.

The antibodies may be coupled to the support using any of the coupling reagents typically used in immunoaffinity chromatography, including cyanogen bromide. After coupling the antibody to the support, the support is contacted with a sample which contains a target polypeptide whose isolation, purification or enrichment is desired. The target polypeptide may be a polypeptide encoded by the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids or the target polypeptide may be one of the EST-related polypeptides, fragments of EST-related polypeptides, positional segments of EST-related polypeptides, or fragments of positional segments of EST-related polypeptides. The target polypeptides may also be polypeptides which have been linked to the polypeptides encoded by the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids or the target polypeptides may be polypeptides which have been linked to EST-related polypeptides, fragments of EST-related polypeptides, positional segments of EST-related polypeptides, or fragments of positional segments of EST-related polypeptides using the fusion vectors described above.

Preferably, the sample is placed in contact with the support for a sufficient amount of time and under appropriate conditions to allow at least 50% of the target polypeptide to specifically bind to the antibody coupled to the support.

Thereafter, the support is washed with an appropriate wash solution to remove polypeptides which have non-specifically adhered to the support. The wash solution may be any of those typically employed in immunoaffinity chromatography, including PBS, Tris-lithium chloride buffer (0.1M lysine base and 0.5M lithium chloride, pH 8.0), Tris-hydrochloride buffer (0.05M Tris-hydrochloride, pH 8.0), or Tris/Triton/NaCl buffer (50mM Tris.cl, pH 8.0 or 9.0, 0.1% Triton X-100, and 0.5MNaCl).

For each of the five genomic DNA libraries, a first PCR reaction is performed according to the manufacturer's instructions (which are incorporated herein by reference) using an outer adapter primer provided in the kit and an outer gene specific primer. The gene specific primer should be selected to be specific for 5' EST of interest and should have a melting temperature, length, and location in the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids which is consistent with its use in PCR reactions. Each first PCR reaction contains 5ng of genomic DNA, 5 µl of 10X Tth reaction buffer, 0.2 mM of each dNTP, 0.2 µM each of outer adapter primer and outer gene specific primer, 1.1 mM of $Mg(OAc)_2$, and 1 µl of the Tth polymerase 50X mix in a total volume of 50 µl. The reaction cycle for the first PCR reaction is as follows: 1 min at 94°C / 2 sec at 94°C, 3 min at 72°C (7 cycles) / 2 sec at 94°C, 3 min at 67°C (32 cycles) / 5 min at 67°C.

The product of the first PCR reaction is diluted and used as a template for a second PCR reaction according to the manufacturer's instructions using a pair of nested primers which are located internally on the amplicon resulting from the first PCR reaction. For example, 5 µl of the reaction product of the first PCR reaction mixture may be diluted 180 times. Reactions are made in a 50 µl volume having a composition identical to that of the first PCR reaction except the nested primers are used. The first nested primer is specific for the adapter, and is provided with the GenomeWalker™ kit. The second nested primer is specific for the particular EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids for which the promoter is to be cloned and should have a melting temperature, length, and location in the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids which is consistent with its use in PCR reactions. The reaction parameters of the second PCR reaction are as follows: 1 min at 94°C / 2 sec at 94°C, 3 min at 72°C (6 cycles) / 2 sec at 94°C, 3 min at 67°C (25 cycles) / 5 min at - 67°C. The product of the second PCR reaction is purified, cloned, and sequenced using standard techniques.

Alternatively, two or more human genomic DNA libraries can be constructed by using two or more restriction enzymes. The digested genomic DNA is cloned into vectors which can be converted into single stranded, circular, or linear DNA. A biotinylated oligonucleotide comprising at least 15 nucleotides from the EST-related nucleic acids,

grow on the selective media contain genes encoding proteins which bind the target sequence. The inserts in the genes encoding the fusion proteins are further characterized by sequencing. In addition, the inserts may be inserted into expression vectors or *in vitro* transcription vectors. Binding of the polypeptides encoded by the inserts to the promoter DNA may be confirmed by techniques familiar to those skilled in the art, such as gel shift analysis or DNAse protection analysis.

VII. Use of EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids in Gene Therapy

The present invention also comprises the use of EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids in gene therapy strategies, including antisense and triple helix strategies as described in Examples 56 and 57 below. In antisense approaches, nucleic acid sequences complementary to an mRNA are hybridized to the mRNA intracellularly, thereby blocking the expression of the protein encoded by the mRNA. The antisense sequences may prevent gene expression through a variety of mechanisms. For example, the antisense sequences may inhibit the ability of ribosomes to translate the mRNA. Alternatively, the antisense sequences may block transport of the mRNA from the nucleus to the cytoplasm, thereby limiting the amount of mRNA available for translation. Another mechanism through which antisense sequences may inhibit gene expression is by interfering with mRNA splicing. In yet another strategy, the antisense nucleic acid may be incorporated in a ribozyme capable of specifically cleaving the target mRNA.

EXAMPLE 56

Preparation and Use of Antisense Oligonucleotides

The antisense nucleic acid molecules to be used in gene therapy may be either DNA or RNA sequences. They may comprise a sequence complementary to the sequence of the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids. The antisense nucleic acids should have a length and melting temperature sufficient to permit formation of an intracellular duplex with sufficient stability to inhibit the expression of the mRNA in the

duplex. Strategies for designing antisense nucleic acids suitable for use in gene therapy are disclosed in Green *et al.*, *Ann. Rev. Biochem.* **55**:569-597 (1986) and Izant and Weintraub, *Cell* **36**:1007-1015 (1984), which are hereby incorporated by reference.

In some strategies, antisense molecules are obtained from a nucleotide sequence encoding a protein by reversing the orientation of the coding region with respect to a promoter so as to transcribe the opposite strand from that which is normally transcribed in the cell. The antisense molecules may be transcribed using *in vitro* transcription systems such as those which employ T7 or SP6 polymerase to generate the transcript. Another approach involves transcription of the antisense nucleic acids *in vivo* by operably linking DNA containing the antisense sequence to a promoter in an expression vector.

Alternatively, oligonucleotides which are complementary to the strand normally transcribed in the cell may be synthesized *in vitro*. Thus, the antisense nucleic acids are complementary to the corresponding mRNA and are capable of hybridizing to the mRNA to create a duplex. In some embodiments, the antisense sequences may contain modified sugar phosphate backbones to increase stability and make them less sensitive to RNase activity. Examples of modifications suitable for use in antisense strategies are described by Rossi *et al.*, *Pharmacol. Ther.* **50(2)**:245-254, (1991) which is hereby incorporated by reference.

Various types of antisense oligonucleotides complementary to the sequence of the EST-related nucleic acids, positional segments of EST-related nucleic acids or fragments of positional segments of EST-related nucleic acids may be used. In one preferred embodiment, stable and semi-stable antisense oligonucleotides described in International Application No. PCT WO94/23026, hereby incorporated by reference, are used. In these molecules, the 3′ end or both the 3′ and 5’ ends are engaged in intramolecular hydrogen bonding between complementary base pairs. These molecules are better able to withstand exonuclease attacks and exhibit increased stability compared to conventional antisense oligonucleotides.

In another preferred embodiment, the antisense oligodeoxynucleotides against herpes simplex virus types 1 and 2 described in International Application No. WO 95/04141, hereby incorporated by reference, are used.

In yet another preferred embodiment, the covalently cross-linked antisense oligonucleotides described in International Application No. WO 96/31523, hereby

SEQ ID NOs: 24-4100 and 8178-36681 or the polypeptide codes of SEQ ID NOs: 4101-8177. The following list is intended not to limit the invention but to provide guidance to programs and databases which are useful with the nucleic acid codes of SEQ ID NOs: 24-4100 and 8178-36681 or the polypeptide codes of SEQ ID NOs: 4101-8177. The programs and databases which may be used include, but are not limited to: MacPattern (EMBL), DiscoveryBase (Molecular Applications Group), GeneMine (Molecular Applications Group), Look (Molecular Applications Group), MacLook (Molecular Applications Group), BLAST and BLAST2 (NCBI), BLASTN and BLASTX (Altschul et al, *J. Mol. Biol.* 215: 403 (1990)), FASTA (Pearson and Lipman, *Proc. Natl. Acad. Sci. USA,* 85: 2444 (1988)), FASTDB (Brutlag et al. Comp. App. Biosci. 6:237-245, 1990), Catalyst (Molecular Simulations Inc.), Catalyst/SHAPE (Molecular Simulations Inc.), Cerius^2.DBAccess (Molecular Simulations Inc.), HypoGen (Molecular Simulations Inc.), Insight II, (Molecular Simulations Inc.), Discover (Molecular Simulations Inc.), CHARMm (Molecular Simulations Inc.), Felix (Molecular Simulations Inc.), DelPhi, (Molecular Simulations Inc.), QuanteMM, (Molecular Simulations Inc.), Homology (Molecular Simulations Inc.), Modeler (Molecular Simulations Inc.), ISIS (Molecular Simulations Inc.), Quanta/Protein Design (Molecular Simulations Inc.), WebLab (Molecular Simulations Inc.), WebLab Diversity Explorer (Molecular Simulations Inc.), Gene Explorer (Molecular Simulations Inc.), SeqFold (Molecular Simulations Inc.), the EMBL/Swissprotein database, the MDL Available Chemicals Directory database, the MDL Drug Data Report data base, the Comprehensive Medicinal Chemistry database, Derwents's World Drug Index database, the BioByteMasterFile database, the Genbank database, and the Genseqn database. Many other programs and data bases would be apparent to one of skill in the art given the present disclosure.

Motifs which may be detected using the above programs include sequences encoding leucine zippers, helix-turn-helix motifs, glycosylation sites, ubiquitination sites, alpha helices, and beta sheets, signal sequences encoding signal peptides which direct the secretion of the encoded proteins, sequences implicated in transcription regulation such as homeoboxes, acidic stretches, enzymatic active sites, substrate binding sites, and enzymatic cleavage sites.

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 6,783,961 B1

DATED : August 31, 2004

INVENTORS : Jean-Baptiste Dumas Milne Edwards, Aymeric Duclert, Jean-Yves Giordano

It is certified that errors appear in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 1,
Line 27, "kilobases (cb)" should read -- kilobases (kb) --.

Column 2,
Line 31, "fraes" should read -- frames --.
Line 31, "Biotech" should read -- Biochem --.

Column 15,
Line 15, "code of SEQ D" should read -- code of SEQ ID --.

Column 16,
Line 24, "The "a" in the" should read -- The "/" in the --.
Line 46, "The "P" in the" should read -- "The "/" in the --.

Column 18,
Line 8, "fingal" should read -- fungal --.
Line 55, "(Strajagene)" should read -- (Stratagene) --.

Column 21,
Line 33, "subunit a and" should read -- subunit α and --.

MAILING ADDRESS OF SENDER:
Saliwanchik, Lloyd & Saliwanchik
P.O. Box 142950
Gainesville, FL 32614-2950

PATENT NO. 6,783,961
No. of add'l. copies
@ 30¢ per page

FORM PTO-1050 (REV. 3-75) UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 6,783,961 B1

DATED : August 31, 2004

INVENTORS : Jean-Baptiste Dumas Milne Edwards, Aymeric Duclert, Jean-Yves Giordano

It is certified that errors appear in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 22,
Line 22, "convex" should read -- connvex --.
Line 25, "convex" should read -- connvex --.

Column 23,
Line 58, "(SEQ IE NOs:3721-4100)" should read -- (SEQ ID NOs:3721-4100) --.

Column 27,
Line 33, "biotin-UJT" should read -- biotin-UTP --.
Line 34, "DIG-UITP" should read -- DIG-UTP --.

Column 28,
Line 49, "270:467470" should read -- 270:467-470 --.

Column 29.
Line 11, "Pictu" should read -- Pietu --.

Column 31,
Line 8, "bioinfomatics" should read -- bioinformatics --.

Column 32,
Line 29, "if the fall" should read -- if the full --.

MAILING ADDRESS OF SENDER:
Saliwanchik, Lloyd & Saliwanchik
P.O. Box 142950
Gainesville, FL 32614-2950

PATENT NO. 6,783,961
No. of add'l. copies
@ 30¢ per page

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 6,783,961 B1

DATED : August 31, 2004

INVENTORS : Jean-Baptiste Dumas Milne Edwards, Aymeric Duclert, Jean-Yves Giordano

It is certified that errors appear in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 36,
Line 31, "Examples 14" should read -- Examples 1-4 --.

Column 37,
Lines 16-17, "(fraction G+C) 0.63% formamide-(600/N)" should read
-- (fraction G+C)-(0.63% formamide-(600/N) --.
Line 53, "Lower Dewees" should read -- Lower Degrees --.

Column 39,
Line 48, "Cutoff Score, Gap" should read -- Cutoff Score=1, Gap --.

Column 40,
Line 45, "Mismatch Penaltyl," should read -- Mismatch Penalty=1, --.

Column 43,
Lines 8-9, "78-8-3-E6-CL1_1C)" should read -- 78-8-3-E6-CL0_1C) --.
Lines 50-51, "108-004-05-0D10-FLC)" should read -- 108-004-05-0-D10-FLC) --.

Column 49,
Line 15, "DAIG" should read -- DA1G --.

MAILING ADDRESS OF SENDER:
Saliwanchik, Lloyd & Saliwanchik
P.O. Box 142950
Gainesville, FL 32614-2950

PATENT NO. 6,783,961
No. of add'l. copies
@ 30¢ per page

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 6,783,961 B1

DATED : August 31, 2004

INVENTORS : Jean-Baptiste Dumas Milne Edwards, Aymeric Duclert, Jean-Yves Giordano

It is certified that errors appear in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 51,
Line 59, "blocling" should read -- blocking --.

Column 52,
Line 44, "CTLA41g" should read -- CTLA4Ig --.
Line 49, "pp. 846847" should read -- pp. 846-847 --.

Column 53,
Line 60, "IIα chain" should read -- II β chain --.

Column 54,
Line 16, "Hematogoiesis" should read -- Hematopoiesis --.

Column 56,
Line 53, "neuropatuies" should read -- neuropathies --.

Column 57,
Line 58, "inhibin a" should read -- inhibin α --.

Column 61,
Line 45, "arc inserted" should read -- are inserted --.

MAILING ADDRESS OF SENDER:
Saliwanchik, Lloyd & Saliwanchik
P.O. Box 142950
Gainesville, FL 32614-2950

PATENT NO. 6,783,961
No. of add'l. copies
@ 30¢ per page

FORM PTO-1050 (REV. 3-75) UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 6,783,961 B1

DATED : August 31, 2004

INVENTORS : Jean-Baptiste Dumas Milne Edwards, Aymeric Duclert, Jean-Yves Giordano

It is certified that errors appear in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 64,
Line 30, "pug/ml." should read -- µg/ml. --.

Column 65,
Line 13, "semifuantitatively" should read -- semi-quantitatively --.

Column 69,
Line 25, "with 2 using" should read -- with P^{32} using --.
Line 26, "kinase Pharmacia)." should read -- kinase (Pharmacia). --.

Column 70,
Line 44, "Ma ping" should read -- Mapping --.

Column 71,
Line 11, "region are surrounding" should read -- region surrounding --.
Line 14, "14:574584" should read -- 14:574584 --.

Column 73,
Line 35, "Acids Positional" should read -- Acids, Positional --.
Line 57, "posifioinal" should read -- positional --.

MAILING ADDRESS OF SENDER:
Saliwanchik, Lloyd & Saliwanchik
P.O. Box 142950
Gainesville, FL 32614-2950

PATENT NO. 6,783,961
No. of add'l. copies
@ 30¢ per page

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 6,783,961 B1

DATED : August 31, 2004

INVENTORS : Jean-Baptiste Dumas Milne Edwards, Aymeric Duclert, Jean-Yves Giordano

It is certified that errors appear in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 76,
Line 13, "DEAE-Dextrar" should read -- DEAE-Dextran --.
Lines 63-64, "fragments a of" should read -- fragments of --.

Column 77,
Line 61, "semiualitative" should read -- semi-qualitative --.

Column 78,
Line 15, "rhodailne" should read -- rhodamine --.
Line 25, "example 1251" should read -- example ^{125}I --

Column 80,
Line 53, "CLAB" should read -- CL-4B --.

Column 82,
Line 18, "DNTP" should read -- dNTP --.

Column 86,
Line 28, "gel shill" should read -- gel shift --.

MAILING ADDRESS OF SENDER:
Saliwanchik, Lloyd & Saliwanchik
P.O. Box 142950
Gainesville, FL 32614-2950

PATENT NO. 6,783,961
No. of add'l. copies
@ 30¢ per page

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 6,783,961 B1

DATED : August 31, 2004

INVENTORS : Jean-Baptiste Dumas Milne Edwards, Aymeric Duclert, Jean-Yves Giordano

It is certified that errors appear in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 87,
Lines 20-21, "acids are A complementary" should read -- acids are complementary --.

Column 97,
Line 14, "EMBUSwissprotein" should read – EMBL/Swissprotein --.

MAILING ADDRESS OF SENDER:
Saliwanchik, Lloyd & Saliwanchik
P.O. Box 142950
Gainesville, FL 32614-2950

PATENT NO. 6,783,961

No. of add'l. copies

@ 30¢ per page

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 6,783,961 B1

DATED : August 31, 2004

INVENTORS : Jean-Baptiste Dumas Milne Edwards, Aymeric Duclert, Jean-Yves Giordano

It is certified that errors appear in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 1,
Line 27, "kilobases (cb)" should read -- kilobases (kb) --.

Column 2,
Line 31, "fraes" should read -- frames --.
Line 31, "Biotech" should read -- Biochem --.

Column 15,
Line 15, "code of SEQ D" should read -- code of SEQ ID --.

Column 16,
Line 24, "The "a" in the" should read -- The "/" in the --.
Line 46, "The "P" in the" should read -- "The "/" in the --.

Column 18,
Line 8, "fingal" should read -- fungal --.
Line 55, "(Strajagene)" should read -- (Stratagene) --.

Column 21,
Line 33, "subunit a and" should read -- subunit α and --.

MAILING ADDRESS OF SENDER:
Saliwanchik, Lloyd & Saliwanchik
P.O. Box 142950
Gainesville, FL 32614-2950

PATENT NO. 6,783,961

No. of add'l. copies
@ 30¢ per page

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 6,783,961 B1

DATED : August 31, 2004

INVENTORS : Jean-Baptiste Dumas Milne Edwards, Aymeric Duclert, Jean-Yves Giordano

It is certified that errors appear in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 22,
Line 22, "convex" should read -- connvex --.
Line 25, "convex" should read -- connvex --.

Column 23,
Line 58, "(SEQ IE NOs:3721-4100)" should read -- (SEQ ID NOs:3721-4100) --.

Column 27,
Line 33, "biotin-UJT" should read -- biotin-UTP --.
Line 34, "DIG-UITP" should read -- DIG-UTP --.

Column 28,
Line 49, "270:467470" should read -- 270:467-470 --.

Column 29.
Line 11, "Pictu" should read -- Pietu --.

Column 31,
Line 8, "bioinfomatics" should read -- bioinformatics --.

Column 32,
Line 29, "if the fall" should read -- if the full --.

MAILING ADDRESS OF SENDER:
Saliwanchik, Lloyd & Saliwanchik
P.O. Box 142950
Gainesville, FL 32614-2950

PATENT NO. 6,783,961
No. of add'l. copies
@ 30¢ per page

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 6,783,961 B1

DATED : August 31, 2004

INVENTORS : Jean-Baptiste Dumas Milne Edwards, Aymeric Duclert, Jean-Yves Giordano

It is certified that errors appear in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 36,
Line 31, "Examples 14" should read -- Examples 1-4 --.

Column 37,
Lines 16-17, "(fraction G+C) 0.63% formamide-(600/N)" should read
-- (fraction G+C)-(0.63% formamide-(600/N) --.
Line 53, "Lower Dewees" should read -- Lower Degrees --.

Column 39,
Line 48, "Cutoff Score, Gap" should read -- Cutoff Score=1, Gap --.

Column 40,
Line 45, "Mismatch Penaltyl," should read -- Mismatch Penalty=1, --.

Column 43,
Lines 8-9, "78-8-3-E6-CL1_1C)" should read -- 78-8-3-E6-CL0_1C) --.
Lines 50-51, "108-004-05-0D10-FLC)" should read -- 108-004-05-0-D10-FLC) --.

Column 49,
Line 15, "DAIG" should read -- DA1G --.

MAILING ADDRESS OF SENDER:
Saliwanchik, Lloyd & Saliwanchik
P.O. Box 142950
Gainesville, FL 32614-2950

PATENT NO. 6,783,961
No. of add'l. copies
@ 30¢ per page

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 6,783,961 B1

DATED : August 31, 2004

INVENTORS : Jean-Baptiste Dumas Milne Edwards, Aymeric Duclert, Jean-Yves Giordano

It is certified that errors appear in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 51,
Line 59, "blocling" should read -- blocking --.

Column 52,
Line 44, "CTLA41g" should read -- CTLA4Ig --.
Line 49, "pp. 846847" should read -- pp. 846-847 --.

Column 53,
Line 60, "IIα chain" should read -- II β chain --.

Column 54,
Line 16, "Hematogoiesis" should read -- Hematopoiesis --.

Column 56,
Line 53, "neuropatuies" should read -- neuropathies --.

Column 57,
Line 58, "inhibin a" should read -- inhibin α --.

Column 61,
Line 45, "arc inserted" should read -- are inserted --.

MAILING ADDRESS OF SENDER:
Saliwanchik, Lloyd & Saliwanchik
P.O. Box 142950
Gainesville, FL 32614-2950

PATENT NO. 6,783,961
No. of add'l. copies
@ 30¢ per page

FORM PTO-1050 (REV. 3-75) UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 6,783,961 B1

DATED : August 31, 2004

INVENTORS : Jean-Baptiste Dumas Milne Edwards, Aymeric Duclert, Jean-Yves Giordano

It is certified that errors appear in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 64,
Line 30, "pug/ml." should read -- µg/ml. --.

Column 65,
Line 13, "semifuantitatively" should read -- semi-quantitatively --.

Column 69,
Line 25, "with 2 using" should read -- with P^{32} using --.
Line 26, "kinase Pharmacia)." should read -- kinase (Pharmacia). --.

Column 70,
Line 44, "Ma ping" should read -- Mapping --.

Column 71,
Line 11, "region are surrounding" should read -- region surrounding --.
Line 14, "14:574584" should read -- 14:574584 --.

Column 73,
Line 35, "Acids Positional" should read -- Acids, Positional --.
Line 57, "posifioinal" should read -- positional --.

MAILING ADDRESS OF SENDER:
Saliwanchik, Lloyd & Saliwanchik
P.O. Box 142950
Gainesville, FL 32614-2950

PATENT NO. 6,783,961
No. of add'l. copies
@ 30¢ per page

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 6,783,961 B1

DATED : August 31, 2004

INVENTORS : Jean-Baptiste Dumas Milne Edwards, Aymeric Duclert, Jean-Yves Giordano

It is certified that errors appear in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 76,
Line 13, "DEAE-Dextrar" should read -- DEAE-Dextran --.
Lines 63-64, "fragments a of" should read -- fragments of --.

Column 77,
Line 61, "semiualitative" should read -- semi-qualitative --.

Column 78,
Line 15, "rhodailne" should read -- rhodamine --.
Line 25, "example 1251" should read -- example ^{125}I --

Column 80,
Line 53, "CLAB" should read -- CL-4B --.

Column 82,
Line 18, "DNTP" should read -- dNTP --.

Column 86,
Line 28, "gel shill" should read -- gel shift --.

MAILING ADDRESS OF SENDER:
Saliwanchik, Lloyd & Saliwanchik
P.O. Box 142950
Gainesville, FL 32614-2950

PATENT NO. 6,783,961
No. of add'l. copies
@ 30¢ per page

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 6,783,961 B1

DATED : August 31, 2004

INVENTORS : Jean-Baptiste Dumas Milne Edwards, Aymeric Duclert, Jean-Yves Giordano

It is certified that errors appear in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 87,
Lines 20-21, "acids are A complementary" should read -- acids are complementary --.

Column 97,
Line 14, "EMBUSwissprotein" should read – EMBL/Swissprotein --.

MAILING ADDRESS OF SENDER:
Saliwanchik, Lloyd & Saliwanchik
P.O. Box 142950
Gainesville, FL 32614-2950

PATENT NO. 6,783,961
No. of add'l. copies
@ 30¢ per page

FORM PTO-1050 (REV. 3-75) UNITED STATES PATENT AND TRADEMARK OFFICE

FEB 2 8 2005